Investigating the Influence of Language Evolution on Perception of a Continuous Meaning Space

9994503

MSc in Evolution of Language and Cognition

The University of Edinburgh

August 2010

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS OF THE DEGREE OF MASTER OF SCIENCE

Declaration

I have read and understood the University of Edinburgh guidelines on plagiarism and declare that that this written dissertation is all my own work except where I indicated otherwise by proper use of quotes and references.

Kate Swoboda

Abstract

There has been a recent resurgence in psycholinguistic experimental studies testing linguistic relativity – the view, associated with Benjamin Lee Whorf (1956), that the language we speak influences our perception and understanding of the environment. Furthermore, recent experimental work in evolutionary linguistics in the Iterated Learning Model applied to humans (e.g., Kirby, Cornish, & Smith, 2008) has proved to be successful at explaining how language transmission can shape its properties. In light of these findings, the research presented here unprecedentedly embarks on testing linguistic relativity from an evolutionary perspective. It is demonstrated that language transmission of two qualitatively different languages evolved in the experiment by Matthews, Kirby, & Cornish (in prep.) – one that promotes the distinction between rotated and unrotated shapes and the other that does not – influences perception of these shapes. This finding suggests that language evolution shapes the conceptual system (semantics) in the speakers' minds by propagating the ways the world should be perceived and interpreted within a given speech community.

Acknowledgements

First of all, I would like to thank my supervisors, Julia Simner and Simon Kirby, for their invaluable advice and help with this project, as well as for their support during my moments of doubt.

I would also like to thank Cristina Matthews for allowing me to test the idea that has grown out of her own Master's Dissertation last year and for some inspiring discussions with her.

My gratitude also goes to Richard Shillcock who helped me with a crucial part of my experimental design and to LEC PhD students, Hannah, Gareth, Sean and Keelin, who were always ready to share with me their academic experiences.

Most importantly, however, I wish to thank my beloved who was always there to believe in me.

Table of Contents

Chapter 1 Introduction

> *‘We dissect nature along lines laid down by our native language. (...)We cut nature up, organize it into concepts, and ascribe significances as we do, largely because we are parties to an agreement to organize it in this way—an agreement that holds throughout our speech community and is codified in the patterns of our language.’* (Whorf, 1956:213)

Whether and to what extent our native language influences the way we perceive, understand and think about the world, as long time ago suggested by Benjamin Lee Whorf, has fascinated researchers and laypeople alike.

Each of us has probably wondered at some point whether cultural differences, including language, can influence our perception and thinking to the extent where it would be impossible to communicate about certain issues.

Although answers to this question will largely remain in the domain of our experiences, the interest in the relationship between the linguistic system and other cognitive domains, referred to as linguistic relativity, has recently enjoyed resurgence in psychological and psycholinguistic research.

Numerous studies in a number of different cognitive domains, such as, for example, spatial cognition (e.g., Levinson, 1996), number cognition (e.g., Gordon, 2004), shape (e.g., Kuo & Sera, 2008) and colour (e.g., Gilbert, Regier, Kay, & Ivry, 2006) have found support for linguistic relativity.

These findings can be interpreted as suggesting that our native language mingles with the make-up of our conceptual system and its semantics. However, their explanatory power has a serious limitation - they do not explain *how* this mingling happens. In other words, explanations that are offered by psychological studies testing linguistic relativity are synchronic and thus do not take into account evolutionary processes that shape language and cognition. In this study we bridge this gap and investigate the influence of language evolution on the evolution of semantics.

An evolutionary approach to linguistic relativity is motivated by a recently growing body of evidence that humans, by evolving linguistic categories (labels), have introduced into their lives an immense evolutionary advantage over non-human animals that have not evolved any language-like communication systems. This advantage enables humans to learn quicker (e.g., Lupyan, Rakison, &

McClelland, 2007) and communicate better (e.g., Steels & Belpaeme, 2005) – to name just two skills we excel at due to our ability to use language.

But of what evolutionary advantage could we speak if we found that evolution of different languages causes changes in semantics of conceptual systems in their speakers? The advantage would be enormous. Such a finding would indicate that language evolution shapes semantics in a way that guarantees communicational intelligibility in given speech communities. In other words, the finding that language evolution shapes semantics in a speech community would mean that what people within this community communicate to each other about the world is perceived and understood in the same way by all community members. This would further indicate that they understand each other better; empathise with each other more, and thus are able to maintain good social relationships and overall social integrity within the community.

Apart from theoretical motivations for the current investigation, there is one more, equally important, factor that has prompted us to design the current experiment. This factor refers to findings in the experiment by Matthews et al. (in prep.) that modelled the evolution of categorisation using the Iterated Leaning Model (ILM).

In the ILM language evolution is modelled as a processes referred to as cultural transmission, whereby generations of learners acquire language from each other by means of observational learning (cf. e.g., Kirby & Hurford, 2002).

Matthews et al.'s (in prep.) experiment has resulted in two findings that are crucial to the current study. Firstly, they demonstrated that evolving languages were becoming more structured by means of organising boundaries of linguistic categories in a continuous meaning space. Secondly, they found that some languages organised the meaning space such that they differentiated rotated and unrotated elements by assigning different labels to them, whereas some other languages did not distinguish between rotated and unrotated elements at all because they assigned identical or very similar labels to both of them.

These two findings constitute a foundation of our research as we hypothesise that the evolution of two languages (L1 and L2), chosen from final evolutionary generations in Matthews et al.'s experiment, influences the way semantics (i.e., the meaning space) is shaped in the minds of speakers of L1 and L2.

In order to investigate the influence of the evolution of L1 and L2 on the evolution of semantics we will measure, in a similarity judgement task, cross-linguistic changes in perception of the meaning space after learning the languages. We will test whether perception of the meaning space differs between L1 and L2 when its elements lie within a category in L1 and across two categories in L2

(Hypothesis 1). Furthermore, we will also test whether perception of rotated and unrotated shapes differs between L1 and L2 which differentiate and do not differentiate rotation respectively (Hypothesis 2).

The last question that needs to be addressed is this: why would cross-linguistic differences in perception of the meaning space tell us anything about changes in semantics?

It is commonly assumed in linguistics that symbols, such as linguistic labels, have semantics (e.g., Chomsky, 1995). However, very often it is unclear what kind of representations semantics refers to and how it is connected to objects in the real world. This situation, referred to as the symbol grounding problem (Harnad, 1990), leads to an implausible model of cognition in which symbols refer to other symbols which refer yet to other symbols, thus resulting in an unrealistic, disconnected from the reality, view of semantics (cf. Searle, 1982 for a Chinese Room argument).

In the present study, however, the symbol grounding problem is resolved because it is assumed that linguistic labels are grounded in action and perception (Barsalou, 1999; Harnad, 1990).This means that the semantics of linguistic labels comprises representations of sensorimotor and perceptual experience. Therefore, if we observe changes in perception as measured in a similarity judgement task, we can infer that there have also been some changes in semantics.

Our assumption that semantics of symbols is grounded in perception and action is compatible with the embodied view of cognition which asserts that human cognitive functions such as language and categorisation, among others, incorporate sensorimotor, perceptual and emotional processes (Gallese & Lakoff, 2005; Wilson, 2002).

In the remainder of this dissertation we review the background literature that motivates our hypotheses and methodology. In Chapter 2 we review literature on categorical perception (CP) – a phenomenon ubiquitous in categorisation that is crucial to our specific assumptions with respect to Hypothesis 1. In the same chapter, we also review studies testing linguistic relativity. In Chapter 3 we review studies on adaptiveness of linguistic labels that theoretically motivate our hypotheses by suggesting that the finding that language evolution influences the way we perceive and understand the world would have an adaptive value. Chapter 4 focuses on the motivations for the ‘cultural transmission’ approach to language evolution and reviews findings that together constitute background to understanding the ILM and Matthews et al.’s experiment on evolution of categorisation. Chapter 5 discusses Matthews et al.’s experiment from the perspective of the current study and describes methods and results of the current research. Chapter 6 is devoted to discussion and Chapter 7 to conclusions.

Chapter 2 To Cognise is to categorise

2.1 What is Categorisation?

Categorisation is a fundamental process underlying human (and non-human animal) cognition that allows us to acquire perceptual and abstract knowledge about the environment (Harnad, 1987a; Harnad, 2005).

As suggested by many categorisation researchers (Estes, 1994; Harnad, 1987a; Sloutsky & Fisher, 2004; Smith, 1989) categorisation is based on induction – that is - a process where we can learn something new about an object based on our past experience with similar objects. Thus, induction can be understood as an inference in which if we know that X1 is Z, and that X1 and X2 are similar, we can conclude that X2 is also Z (Sloutsky & Fisher, 2004). In other words, categorisation requires comparing a new object X2 (e.g., a blackbird) to a known instance, X1, of a category Z (e.g., 'bird'), learning that the two instances are similar because they both have wings, and concluding that X2 (i.e., the blackbird) belongs to the same category Z (i.e., 'bird') as X1. It is worth emphasising that X1 and X2 are compared based on similarity, thus making this concept fundamental with respect to categorisation.

2.2 Categorical Perception

In the current study we investigate the perception of a continuous meaning space organised by language evolution into distinct categories. Thus, it is important to have a closer look at the exact mechanism that allows us to represent the continuous physical environment in the form of separated and bounded categories. Categorical perception (henceforth, CP) is such a mechanism. CP allows us to perceive sensory continua discontinuously, chunking perceptual spaces into discrete categories. Furthermore, CP shapes representations of categories in such a way that it increases similarity between representations of members of the same category and decreases similarity between representations of members of different categories (Harnad, 1987a). If we assume that similarity between two objects is represented as distance in a cognitive similarity space between the representations of these objects (Shepard, 1974)[1], then we can say that CP results in within-category similarity space compression and between-category similarity space expansion. As a result discrete and easily discriminable category representations are formed.

CP was first observed in speech (Liberman, Harris, Hoffman, & Griffith, 1957). Liberman and colleagues demonstrated that sounds /ba/ and /da/ which varied systematically along the voice-onset

[1] In §6.3.3 we discuss the relevance of the concept of similarity to our research in more detail.
[2] The Dani were claimed by Rosch Heider and Olivier (1972) to have just two basic colour terms.
[3] Berinmo is a language spoken in Papua New-Guinea.
[4] 'Nol' covers green, blue and purple; 'wor' covers yellow, orange, brown and khaki.

time continuum, were discriminated more accurately when they were in different phonemic categories (i.e., one in /ba/, another in /da/), compared to when they were in the same phonemic category (i.e., both in /ba/, or both in /da/). Similar speech CP effects were also found in prelinguistic infants (e.g., Eimas, Siqueland, Jusczyk, & Vigorito, 1971), as well as in non-human animals (e.g., 1987; Kuhl & Miller, 1975; Snowdon, 1987).

Furthermore, CP effects have also been detected in other domains, such as, for example, colour perception (e.g., Gilbert, Regier, Kay, & Ivry, 2006; Roberson, Davies, & Davidoff, 2000; Winawer et al., 2007; Özgen & Davies, 1998) face perception (e.g., Beale & Keil, 1995; Goldstone, Lippa, & Shiffrin, 2001; Levin & Beale, 2000), and interval perception in musicians (e.g., Burns & Ward, 1978), as well as in neural networks during category learning (e.g., Harnad, Hanson, & Lubin, 1995; Nakisa & Plunkett, 1998).

It is worth noting, however, that, although CP is ubiquitous in categorisation, it is also possible to categorically interpret sensory continua without there being CP at work (cf. Ladd & Morton, 1997). Categorisation without CP effects has been demonstrated, for example, for lexical tones in tone languages (Francis, Ciocca, & Ng, 2003) and for vowels (Stevens, Liberman, Studdert-Kennedy, & Ohman, 1969).

2.3 Universalist versus Relativist Debate

Although on the surface CP appears to solely concern perceptual categorisation, it is crucial to note that since language labels a large number of our perceptual categories, it is also likely to influence CP (Harnad, 1987b). In fact, the influence of language on perception and CP is central to the current experiment.

There are two views on the relation between perception and language in the literature (cf. Regier & Kay, 2009). The first view, referred to as universalist, postulates the existence of universal tendencies with respect to perception of sensory stimuli that, instead of being influenced by languages, influence languages themselves (e.g., Berlin & Kay, 1969; Rosch Heider & Olivier, 1972; Rosch, 1973). In this view, CP effects are considered universal and do not differ across languages. The second view, associated with Whorf (1956) and referred to as linguistic relativity, postulates that our perception of sensory stimuli is influenced by the language we speak (Kay & Kempton, 1984; Roberson, Davidoff, & Shapiro, 2002; Roberson, Davies, & Davidoff, 2000; Winawer et al., 2007; Özgen & Davies, 1998) Accordingly, CP effects are shaped by cross-linguistic differences with respect to linguistic category labelling.

In what follows I focus on demonstrating arguments of both universalists and relativists with respect to the relation between language and perception in the domain of colour. This domain has been chosen due to its representativeness of the debate (cf. Regier & Kay, 2009). Although the domain of visually perceived shape is relevant to the current experiment it will not be reviewed here due to the space limit (cf. e.g., Rosch, (1973) – for a universalist view, and Roberson et al., (2002) – for a linguistic relativist view).

2.3.1 *The Case of Colour*

In 1969, Berlin and Kay proposed that there exist universal focal colours around which labelled colour categories are organised in all languages. To test Berlin and Kay's (1969) influential proposal, Rosch Heider & Olivier (1972) performed a number of experiments that compared colour naming and memory between American-English speakers and the Dugum Dani (henceforth, the Dani), an agricultural population in Irian Jaya, Indonesia.

For example, one of their experiments was claimed to support the universal perception of colour space by showing that there was no difference between English-American speakers and the Dani with respect to how colours were remembered, despite the differences in colour vocabularies of these two populations[2] . Such a result appeared to indicate that colour labels do not influence the memory of colours in either of the participating group. Thus, the Whorf hypothesis was concluded not to be supported.

However, Rosch Heider and Olivier's (1972) studies were reported to have flaws in the design and the interpretation of results (cf. Lucy, 1997; Saunders & van Brakel, 1997). Given this, Roberson et al.'s (2000) study seems to be of particular importance since not only did it fail to replicate Rosch Heider and Olivier's (1972) results, but it also performed a number of new experiments that supported linguistic relativity.

In one of their experiments (Experiment 4), Roberson et al. (2000) investigated perception of colours in the populations of English and Berinmo[3] speakers. They focused on a between-group comparison of CP effects with respect to colours that straddled the green-blue and *nol-wor* colour boundary. Colours on the former boundary are labelled only in English, whereas colours on the latter boundary only in Berinmo[4]. Roberson et al. (2000) used an odd-one-out matching triad task, where participants had to choose two most similar stimuli out of three. The stimuli consisted of Munsell chips which were manipulated such that in each triad (1) all chips were in one category; or (2) two were in one category, and one on a boundary; or (3) two were in one category and one in another. In each triad the distances between chips (in Munsell steps) were equal. Participants' choices in the task were used to calculate CP effects for each language group.

It was predicted that if colour labels influence colour perception, then CP effects should be observed for the *nol-wor* boundary only in Berinmo speakers and for the green-blue boundary only in English

[2] The Dani were claimed by Rosch Heider and Olivier (1972) to have just two basic colour terms.
[3] Berinmo is a language spoken in Papua New-Guinea.
[4] 'Nol' covers green, blue and purple; 'wor' covers yellow, orange, brown and khaki.

speakers. Roberson et al. (2000) found CP effects significantly more often for the green-blue boundary than for the *nol-wor* boundary for English speakers and the reverse for Berinmo speakers. However, although performance of both Berinmo and English speakers was at chance with respect to the colour boundaries not existing in their respective languages, there was no significant difference between performance of Berinmo and English speakers with respect to the green-blue boundary. Roberson et al. (2000) interpreted this finding as a possibility that, perhaps, there was a universal principle governing the categorisation of greens and blues.

In order to exclude this possibility, Berinmo and English speakers were taught to sort Munsell chips into two piles, one with the category boundary set at the green-blue boundary, as in the previous experiment, and another with an arbitrary category boundary dividing the green spectrum into green1 and green2 (Roberson et al. 2000; Experiment 5). Crucial here was the inclusion of the green1-green2 boundary that was arbitrary for both Berinmo and English speakers.

It was predicted that, if greens and blues perception is not governed by some universal perception principles – contrary to what still remained a possibility in Experiment 4 – then, Berinmo speakers would find it as difficult to learn to sort stimuli into green1 and green2 categories as they would find to learn to sort them into blue and green categories. By contrast, English speakers were predicted to find it harder to learn to classify stimuli straddling the green1-green2 boundary than the green-blue-boundary.

Both these predictions were confirmed and Roberson et al. (2000) concluded that colours in Berinmo and English are categorised according to colour terms available in their respective languages[5].

Roberson et al. (2000) is only one among many other studies claiming to support Whorfian hypothesis in the domain of colour. Similar results were found by Kay & Kempton (1984). They found between-category expansion of similarity space close to the blue-green category boundary for English, but no such effect for Tarahumar, a language spoken in Northern Mexico that does not label a distinction between blue and green, but instead names this part of colour spectrum 'siyòname' (Experiment 1).

More recently, Őzgen & Davies (1998) found evidence for linguistic relativity with respect to Turkish which has two basic terms for blue (*lacivert* and *mavi*). In a sorting task, where participating adults were required to sort light-blue and dark-blue stimuli into two piles based on their similarity, Ozgen and Davies found that Turkish speakers were more likely than English speakers, who use just one label for blue, to put stimuli into two separate categories. In a subsequent similarity judgement task, where participants were asked to rate similarity of stimuli pairs, Turkish speakers demonstrated

[5] Findings in Roberson et al.'s (2000) last two experiments (i.e., Experiments 6a-6b) further support the Whorf hypothesis by confirming their previous results, but in a different task – a two alternative forced choice task.

decreased between-category similarity (i.e., stimuli that crossed the *lacivert-mavi* boundary were judged less similar) compared to no such effect for English speakers.

Winawer et al. (2007) found evidence for the Whorf hypothesis with respect to Russian blues. They found that Russian speakers were faster at discrimination of colour stimuli ranging from light to dark blue when stimuli straddled the lexicalised category boundary *siniy-goluboy*, compared to when the stimuli were in the same lexicalised category, *siniy* or *goluboy*. By contrast, English speakers were equally fast in both conditions.

A pioneering study by Gilbert, Regier, Kay, & Ivry (2006) demonstrated that the Whorf effect is present only in the Right Visual Field (RVF). Crucially, this idea is a consequence of the fact that language is mainly located in the Left Hemisphere (LH) and that projections from visual fields are processed contralaterally. Thus, if linguistic relativity is to be supported, it is predicted to involve only the RVF.

In a pre-experimental task participants established the linguistically labelled blue-green boundary. The main experiment involved a visual search task in which participants were instructed to state if the target stimulus was located in the left or the right half of the computer screen. Stimuli consisted of coloured squares surrounding centrally located fixation point. All stimuli were the same colour (distracters) apart from one square (target) which was different. In order to investigate CP effects, the distracters and the target were included in the same category (e.g., containing blues) or in two different categories (e.g., one containing blues and the other greens). The CP effect – that is – an improved discriminability of between-category colours, compared to the within-category discriminability, was found only in the RVF.

Some recent findings that will not be reviewed here due to the space limit reveal a complicated picture suggesting that also universal colour naming constraints should be incorporated into a fully explanatory account of colour categorisation (cf. e.g., Kay & Regier, 2007; Lindsey & Brown, 2006). However, in light of the abundance of evidence in favour of linguistic relativity, the current consensus is that colour perception and categorisation is influenced by the language one speaks and that this influence can be observed in the RVF (cf. Regier & Kay, 2009).

2.4 CP as Affected by Development

Our expectations with respect to the first hypothesis of the current experiment that CP effects will be induced during language transmission are supported by the profusion of studies demonstrating that CP mechanisms are extremely flexible and that they are shaped by language learning. This flexibility can be demonstrated not only when we compare CP effects cross-linguistically, but also when we compare participants speaking the same language but at different developmental stages.

For example, Franklin, Clifford, Williamson, & Davies (2005) have demonstrated that, both Himba and English toddlers show CP for the blue-purple category boundary before they have acquired any colour labels. However, in accordance with linguistic relativity, CP effects for the blue-purple boundary are no longer present in the Himba, after they have acquired their language's colour terms (Roberson, Davidoff, Davies, & Shapiro, 2005).

In addition, Franklin, Drivonikou, Bevis, et al. (2008a) demonstrated that colour CP in prelinguistic 4-6 month old infants is only present in the LVF. Given the findings by Gilbert et al. (2006) that CP effects in adults are found only in the RVF, it is possible that the switch from the LVF to the RVF is a developmental change and takes place due to the influence of acquisition of colour labels. This possibility was tested by Franklin, Drivonikou, Clifford, et al. (2008b). They demonstrated that 32 month old toddlers, who were in the process of learning colour labels (referred to as 'learerns'), demonstrated CP effects in the LFV. By contrast, 46 month olds, who have mastered accurate use and understanding of colour terms (referred to as 'namers'), demonstrated CP effects in the RVF. Notably, the scope of CP effects did not differ between 'learners' and 'namers'.

2.5 CP Effects in Artificial Category Learning

On linguistic relativity view, the same meaning space can be perceived differently by language A and B, if, for example, language A has just one label (category) to refer to it, while language B has two. Similarly, in an artificial category learning experiment the perceptual dimension can be perceived differently relative to the number of labels (artificial categories) that are assigned to it. Given this, we can expect to find similar effects in the current study – that is - shape stimuli are predicted to be perceived differently or identically, depending on whether their similarity would be judged by speakers of languages that labelled them differently or identically.

Between-category expansion and within-category compression (i.e., CP effects) have been demonstrated as a result of artificial category learning in a number of studies using different perceptual stimuli. For example, CP effects have been demonstrated for continua of familiar faces (Beale & Keil, 1995), continua of unfamiliar and inverted faces (Levin & Beale, 2000), or animal body parts and artificial cells (Livingston, Andrews, & Harnad, 1998). However, Goldstone's (1994a) experiment will be described in more detail because it uses geometrical figures as stimuli, as in our experiment .

Goldstone (1994a, Experiment 2) conducted an artificial category learning experiment investigating how perception of stimuli continuously differing along two perceptual dimensions, such as, size and brightness, changes as a result of assigning two labels (A and B) to the stimuli. The stimuli consisted of 16 squares arranged in a matrix 4x4, where horizontally and vertically adjacent squares differed in their size and brightness, respectively. Participants were assigned to three conditions: size and brightness categorisers and a control group. Size categorisers learned with feedback to categorise stimuli organised such that two left columns of the matrix were labelled A and two right columns were labelled B. By contrast, brightness categorisers learned with feedback to categorise stimuli such that two top and bottom rows belonged to categories A and B, respectively.

After learning, all participants performed a same/different discrimination task, where they were presented with a pair of squares, appearing on the screen successively, and required to respond S (for same) or D (for different). The stimuli were pairs of squares that differed in size and brightness for size and brightness categorisers respectively, and that were inside a category (within-category pairs) or across categories (between-category pairs).

Goldstone predicted that size and brightness categorisers would make fewer errors than the control group in their perceptual discriminations of between-category pairs. This process, referred to by Goldstone as *acquired distinctiveness*, can be thought of as a between-category expansion of similarity space. He also predicted that both size and brightness categorisers would make more errors than the control group in their perceptual discriminations of within-category pairs. This process, referred to by Goldstone as *acquired equivalence*, can be considered as a within-category compression of similarity space. Goldstone's results demonstrated *acquired distinctiveness* effect for both size and brightness categorisers, but no *acquired equivalence* effect.

The ideas of *acquired distinctiveness* and *acquired equivalence* serve as a motivation for our first hypothesis suggesting differential perception of shape pairs depending on whether they are labelled identically or differently.

Chapter 3 Adaptive role of language labels

3.1 Human versus non-human animal communication systems

There is a great qualitative gap between animal and human vocal communication systems. An animal communication system uses largely innate object-signal associations only[6] (e.g., vervet monkeys use three types of distinct alarm calls as a response to the appearance of an eagle, a leopard or a snake (Seyfarth, Cheney, & Marler, 1980)). By contrast, language uses learnt and largely arbitrary associations between objects and symbols, as well as symbols and other symbols[7].

Although it has been demonstrated that non-human animals in captivity are able to learn to understand and produce single symbols and language-like symbolic sequences (e.g., bottlenose dolphins – Herman (1984); sea lions – Schusterman & Gisiner (1988); a chimpanzee named Nim – Terrace, Petitto, Sanders, & Bever (1979); a bonobo named Kanzi – Savage-Rumbaugh et al. (1993)), they do not seem to be able to use these learning abilities in the wild. Thus, the observed non-human animal linguistic abilities seem to be due to captive enculturation (cf. Bering, 2004).

The ability of humans to use symbols, including language labels, has a great evolutionary advantage over non-human animals, which lack this skill, because it increases availability of strategies that can be used during learning and communication, thus increasing humans' overall fitness.

This chapter concentrates on demonstrating that having linguistic labels is adaptive. In other words, we will attempt to demonstrate that linguistic labels allow humans to be better category learners and communicators. Furthermore, the evidence demonstrated here will serve to motivate our assumption that a crucial link exists between categorical perception and the evolution of language labels, and hence language.

3.2 Sensorimotor Toil versus Symbolic Theft

Cangelosi, Greco, & Harnad (2000), using a three-layer feed-forward neural network, modelled two types of category learning. The first type, 'sensorimotor toil', refers to learning new categories via trial-and-error with feedback, basing the learner's knowledge on her sensorimotor and perceptual experiences. The second type, 'symbolic theft', refers to learning new categories with aid of linguistic labels without the need to ground all categories in experience.

In their model, the neural network learned via 'sensorimotor toil' names of perceptual prototypical categories. It also learned via 'symbolic theft' to associate these grounded category names with novel 'higher order' labels - *symmetric* and *asymmetric*. During testing, it was demonstrated that, when

[6] An obvious counterexample is animal song that is largely learnt (cf. Merker & Okanoya, 2007)

[7] We use Deacon's (1997) terminology here, where 'symbol' can refer to words or phrases.

presented with visual input, the network was capable of producing 'high order' categories (i.e., *symmetric* or *asymmetric*) with 80% success rate.

Assuming that Cangelosi et al's (2000) model is a plausible model of the first stages of language acquisition, it can be concluded that its performance during testing demonstrates that 'symbolic theft' can indeed be a very successful category acquisition strategy.

3.3 Advantage of Learning via Symbolic Theft in a Miniature Evolutionary Scenario

Cangelosi et al.'s (2000) model demonstrates that 'symbolic theft' is a plausible strategy of efficient category acquisition. However, Cangelosi & Harnad (2002) emphasise the advantage of this strategy even more by demonstrating that 'symbolic theft' is more adaptive than 'sensorimotor toil' in a miniature evolutionary scenario.

They compare performance of artificial agents foraging in a simulated environment (cf. Parisi, 1997) that have to learn to distinguish four types of mushrooms: those with feature A or B must be eaten or marked respectively, those with features AB must be eaten, marked and returned to, and those with features C, D or E must be ignored.

During the first set of 200 simulations foragers learn via 'sensorimotor toil' and their population is subjected to selection and reproduction, using a genetic algorithm (Goldberg, 1989). After genetic evolution, the foragers are divided into two groups (i.e., 'toilers' and 'thieves') in which they learn in two stages.

Firstly, both 'toilers' and 'thieves' learn via 'sensorimotor toil' to eat and mark mushrooms A and B, and to produce an appropriate call (i.e., EAT and MARK, respectively). However, they do not learn to respond to 'return' mushrooms.

In the second stage, 'toilers' learn to return to mushrooms AB and to produce the appropriate call (RETURN) via reception of mushrooms' visual features as input. By contrast, when 'thieves' learn the action of returning and the corresponding call (RETURN), they are fed only the call (RETURN) as input. In other words, instead of learning to respond appropriately based on the experience with perceptual features of mushrooms AB, 'thieves' learn from hearing their label. Because this label is indirectly grounded in the experience of mushrooms A and B separately, they are able to understand it.

In order to establish which strategy had an evolutionary advantage, the number of returned mushrooms was compared between 'toilers' and 'thieves'. Given that 'thieves' returned more mushrooms than 'toilers', it was concluded that 'symbolic theft' is a more adaptive label learning strategy than 'sensorimotor toil'.

3.4 Other Experiments Demonstrating Adaptive Advantage of Linguistic Labels

There is evidence in child language development literature that children discriminate labelled objects better than unlabelled objects (Xu, 2002). Furthermore, it has been demonstrated that redundant linguistic labels that were correlated with categories of solid and non-solid things facilitated learning of these categories in two-year old children (Yoshida & Smith, 2005).

In addition, Lupyan, Rakison, & McClelland (2007) demonstrate that participants learn to categorise stimuli (pictures of aliens) more quickly when stimuli are labelled with written (Experiment 1) or auditory (Experiment 2) non-words (*leebish* and *grecious*), compared to when they have no labels, or when the label is a proposition and not a single word (Experiment 2).

What is crucial in Lupyan et al.'s (2007) experiment is that stimuli could also be categorised without labels, according to perceptually salient differences in two perceptual dimensions. Yet, the addition of labels that did not supply any information about stimuli, resulted in accuracy of 80% in categorisation after only 30 training trials, as opposed to 72 training trials in the group where labels were absent.

The result that people better categorise novel stimuli with labels than stimuli without labels could be due to the finding that conceptual knowledge, to which labels contribute, facilitates visual processing on-line by means of supplying top-down feedback (Lupyan, 2008; Lupyan, Thompson-Schill, & Swingley, 2010).

Steels & Belpaeme (2005) , using simulations with artificial agents playing perceptual discrimination and guessing games, also emphasise the adaptive role of language labels (cf. Belpaeme & Bleys, 2005 for a similar account).

They demonstrate that, while learning to categorise the colour space according to statistical distributions of colours from real-world colour samples[8], as opposed to learning from a sample of random data, helps agents to form categories that are shared in a population to a certain degree, it does not result in a complete sharing of the categories that would be required for efficient communication.

Furthermore, Steels and Belpaeme suggest that the best way to guarantee the acquisition of a shared colour category system in a population is to learn categories coupled with language labels. They support this point by showing, in a number of simulations, that individualistic learning via a perceptual discrimination game does not result in a perfectly shared system of colour categories, as well as that genetic evolution requires too long a time to evolve such a system to be an ecologically plausible solution (it would take about 400 years to evolve a shared colour category repertoire in a population of 10 humans, assuming a new generation appeared every 20 years). Furthermore, they

[8] Real-world colour stimuli were obtained from photographs of natural and urban environments.

demonstrate that the acquisition of shared colour categories in a language guessing game, whereby the hearer learns to identify a colour category by means of hearing a name associated with it from the speaker, results in a shared colour category system that enables successful communication. The adaptive role of language labels in the acquisition of a shared colour category system in a language guessing game seems even more impressive if we take into account that agents learned to categorise the colour space from stimuli without any realistic statistical distributions of colours.

Since, as we have seen, language labels have an adaptive role by enabling their users to learn and induce new categories very effectively, it appears reasonable to suggest, as we do here, that language and categorical perception co-evolved, one bound and affected by the other. This suggestion motivates our hypothesis that language evolution causes changes in semantics of conceptual systems in the speakers because these changes lead to adaptive behaviour. We return to this point in §6.4.

Chapter 4 Language evolution

4.1 Language complexity: compositionality

All animal communication systems are structured to certain extent. Some systems, such as bird song (e.g., Kroodsma & Parker, 1977) or the song of humpback whales (Payne, 2000) can even be very complex structurally. Furthermore, symbolic systems can be taught to animals in captivity (cf. § 3.1). Even more impressively a number of different species have been demonstrated to have quite advanced aural pattern recognition abilities (e.g., cotton-top tamarins – Fitch & Hauser (2004); European starlings - Gentner, Fenn, Margoliash, & Nusbaum (2006). Yet none of the non-human animal communication systems are compositional (Kirby & Hurford, 2002) - that is – no non-human communication systems consist of signals where 'the meaning of a signal is some function of the meaning of the parts of that signal and the way in which they are put together' (p.128).

Compositionality is a universal feature of all human languages and together with other universals they comprise these language properties that many linguists have long sought to explain. However, only recently the mainstream approach of linguistic enquiry changed from synchronic (e.g., Chomsky, 1995) to diachronic and evolutionary[9] (e.g., Bickerton, 1990; Croft, 2000; Kirby, in prep.; 2000; 2001; Wray, 1998; Tallerman 2007), thus leading to an explanation of compositionality.

In the remainder of this chapter I will briefly explain reasons for approaching language as a system that evolved during cultural transmission, rather than natural selection. Then, I will describe background findings in computer modelling that allowed researchers to explain some aspects of language acquisition and some language universals. All this background knowledge will lead us to the introduction of the Iterated Learning Model (henceforth, the ILM) and the explanation of compositionality. The literature reviewed in this chapter will help us to better contextualise the experiment by Matthews et al. (in prep.) that demonstrates the evolution of categorisation within the ILM and that constitutes a starting point of our current investigation. Matthews et al.'s study will be discussed in Chapter 5 in more detail.

[9] One of the most frequently debated issues in evolutionary linguistics concerns the debate over characteristics of the first form of language, referred to as protolanguage (e.g., Bickerton, 1990; Tallerman, 2007; Wray, 1998). Although not without critique (e.g., Tallerman, 2007), Wray's (1998) proposal of holistic protolanguage, which over evolutionary time changed into a compositional system, is in line with computer simulations (e.g., Kirby,2001) and experiments with human participants (e.g., Kirby, Cornish, & Smith, 2008) that have been recently revolutionising the field of evolutionary linguistics.

4.2 Natural selection versus cultural transmission

In a seminal paper, Pinker & Bloom (1990) claimed that 'the only way to explain the origin of such abilities as language is through the theory of natural selection'. This means that language and its structural properties should be treated on a par with such functional systems as, for example, the vertebrate eye that has biologically adapted to the requirements imposed by the environment that its bearers inhabit (Lamb, Collin, & Pugh, 2007). However, the idea that a quickly changing linguistic environment could allow for the evolution of language genes has been argued not to be a feasible scenario because natural selection needs a stable environment to engender genetic adaptations (Chater, Reali, & Christiansen, 2009; Christiansen & Chater, 2008).

An alternative view of the evolution of structural properties of language assumes that they are products of cultural transmission (or cultural evolution). Arguably, the most important feature of this approach, that appears to be uniquely human (Price, Caldwell, & Whiten, 2010), is the cumulativeness of behaviours that result from cultural transmission (Boyd & Richerson, 1996; Tomasello, 1999). Cumulative culture is usually present in populations of social learners where 'the accumulation of modifications over time result[s] in innovation that no individual could have discovered on his or her own' (Price et al., 2010: 27). Therefore, cumulative culture results in behaviours that are much more adaptive than the ones acquired individually (Boyd & Richerson, 1996).

In line with this thinking, Kirby & Hurford (2002) suggest that although biological evolution equipped humans with learning mechanisms needed to learn language, it was cultural evolution that led to the emergence of languages that exist today and their universal features.

4.3 Emergent behaviour and language acquisition as the problem of induction

A crucial step in computer modelling of language was the demonstration that linguistic rule-like behaviour can emerge as a result of a computer model's learning, according to a certain learning algorithm. This meant that linguistic behaviour did not have to be explained by the postulation of some inherent explicit rules (such as, for example, those required by Chomsky's (e.g., 2002) concepts of the Language Acquisition Device (LAD) or Universal Grammar (UG)), predefining it before it actually emerged. In fact, it is true that emergent linguistic behaviours in computer models are constrained by the internal architecture of the model and the learning algorithm used, but nowhere in such models are there rules explicitly built into their architecture.

For example, McClelland & Rumelhart (1986) demonstrated, using a simple pattern associator neural network, that the U-shaped acquisition profile of English past-tense forms can result from the network's generalisations over the input data. Although their model has been heavily criticised for the lack of explanation of detailed linguistic facts of English past tense (Pinker & Prince, 1989), it presents a significant step towards an understanding of linguistic behaviour as emergent - that is – learnt and cultural, rather than predetermined by the LAD or UG.

Computer simulations also have helped to address another issue that has troubled linguists for decades, namely, the logical problem of language acquisition (or the poverty of stimulus argument). According to Chomsky, linguistic input that a child gets during language acquisition is ‘too impoverished to motivate the grammatical knowledge that adult speakers invariably possess’ (2002:7). Given this, the argument of the poverty of stimulus seems to motivate the postulation of UG which can supply innate grammatical properties lacking in linguistic input received by a child.

However, in addition to theoretical arguments suggesting that no adequate empirical data has been supplied in favour of Chomsky’s formulation of the poverty of stimulus argument (Pullum & Scholz, 2002), there is evidence from computer simulations and behavioural studies that impoverished learning data is *required* for language acquisition (e.g., Cornish, Tamariz, & Kirby, 2009; Kirby, in press.; Kirby & Hurford, 2002; Kirby, Cornish, & Smith, 2008).

Computer simulations see language learning as an induction problem in which limited language data, constituting a transmission bottleneck, is unavoidable (e.g., Cornish, Tamariz, & Kirby, 2009; Kirby, in press.) and necessary for structured generalisation behaviour to occur (e.g., Elman, 1993; Kirby, 2001; Kirby & Hurford, 2002; Kirby, Cornish, & Smith, 2008). For example, using the ILM (to be explained shortly) Kirby & Hurford (2002) show that only when neural networks are trained on a medium sized set of 50 meaning-signal pairs, as opposed to a very small training set of 20 items and a very large training set of 2000 items, non-random structure emerges.

Furthermore, as noted by Cornish (2010), a data bottleneck is not the only way of limiting language input during learning that can result in the emergence of structure. She emphasises that the most important factor is the presence of imperfect information, while its source is irrelevant. Thus, for example, the limited data can come from the fact that children have limited memory and attention span (Elman, 1993 – simulation 3) or that human adult learners are not perfect learners – that is - even if they get access to the training data comprising the whole language, they cannot make a full use of the provided information due to the constraints imposed by their memory (Cornish, 2010).

4.4 Language as an organism and an explanation of universals

Another crucial step in modelling language learning was the demonstration that emergent linguistic behaviour adapts to the learner’s learning biases, such as for example, cognitive-general sequential learning biases (Christiansen & Ellefson, 2002; Christiansen, Kelly, Shillcock, & Greenfield, under review; Reali & Christiansen, 2009) or a bias against synonyms and homonyms (Smith, 2004). Furthermore, it was argued that some of such learning biases could be innate (Batali, 1998; Smith, 2004). In light of these findings, it was suggested that language can be considered an organism which adapts to pressures coming from its learning environment which is the human mind (Christiansen & Ellefson, 2002; Christiansen & Chater, 2008).

An important contribution to the debate over universals are the findings demonstrating that particular linguistic universals can emerge as a result of language adapting to the learner's learning biases. For example, Christiansen & Devlin (1997), training neural networks with the same internal architecture – hence - the same learning biases, demonstrated that recursively inconsistent languages were harder to learn than recursively consistent languages[10] (cf. Christiansen, 2000 on similar findings with human participants). This finding is interesting because recursively consistent structures are more common in the typological distribution of languages in the world (cf. Dryer, 1992) which suggests a largely universal word order preference for recursive consistency.

4.5 Modelling language evolution via cultural transmission

4.5.1 *The Iterated Learning Model (the ILM)*

In previous sections we have seen that linguistic universals, even without modelling cultural evolutionary processes, can be conceived of as emergent adaptive behaviours which arise as a consequence of fitting into learning biases of the language learner. The leading theme of this section is the idea that if we model language evolution as the process of cultural transmission in the form of iterated learning (Kirby & Hurford, 2002), we can witness, under certain circumstances (e.g., given certain cognitive biases and certain sizes of the transmission bottleneck), the emergence of more learnable and more structured languages (Brighton, Smith, & Kirby, 2005; Cornish, 2010; Cornish, Tamariz, & Kirby, 2009; Kirby, Cornish, & Smith, 2008; Kirby, Smith, & Brighton, 2004). It is important to emphasise that the emergence of such a structure is by no means caused by intentions of artificial or human agents.

The iterated learning is 'a process whereby an individual acquires a behaviour by observing a similar behaviour in another individual who acquired it in the same way' (Kirby et al. 2008: 10681) and has been successfully implemented in the Iterated Learning Model (cf. Mesoudi & Whiten, (2008), for a review of other models of cultural transmission).

The ILM has been extensively used in computational modelling to model language evolution (e.g., Griffiths & Kalish, 2007; Kirby, 2001; Kirby & Hurford, 2002; Kirby, Smith, & Brighton, 2004; Smith, 2002; Smith, 2004). Recently, the ILM has been applied to humans to more realistically model learning biases of human cognition during language evolution (Kirby et al., 2008; Cornish et al., 2009; Cornish, 2010, Matthews et al. in prep.).

[10] Recursively consistent languages are such languages in which the head always precedes or follows the complement. Recursively inconsistent languages, on the other hand, are both head-initial and head-final.

Typically, the ILM comprises a population of individuals arranged into transmission chains. Each member (generation) of the chain has a task of learning an artificial language (the finite set of signal-meaning pairs) and producing an output which serves as the input for the next generation. This cycle is repeated until the desired number of generations has been reached. The transmission of information in each chain mirrors the way information about language structure is transmitted via generalisation over evolutionary time-scale.

4.5.2 *Emergence of compositionality in the ILM*

We started this chapter with a goal of explaining compositionality – the unique feature of natural languages that has been recently described as one of the major transitions in language evolution (Kirby, in press). Now we will demonstrate the circumstance in which it emerges in the ILM.

Although compositionality emerges in a typical computer simulation using the ILM (e.g., Kirby et al., 2004), not all ILM studies with human participants demonstrate emerging compositionality. Instead, systematic underspecification with ambiguity (i.e., the use of same signals for different meanings) appears in some of the experiments (e.g., Experiment 1 in the study by Kirby et al., 2008; an experiment with participants' memory acting as a bottleneck in the study by Cornish, 2010). Only when the learning set was filtered against homonyms did compositionality emerge (Kirby et al., 2008; Experiment 2). Kirby et al., (2008) suggest that, although the filtering was applied artificially, it can be assumed to model the communication pressure of expressivity (i.e., the pressure to express meanings only by means of one-to-one signal-to-meaning mappings, as opposed to one-to-many signal-to-meaning mappings). Furthermore, such an artificial filtering could be treated as an equivalent of an innate learning bias against homonymy that human learners seem to possess (cf. Smith, 2004).

Chapter 5 Current experiment

5.1 Introduction

The ILM methodology we have reviewed in Chapter 4 served as the basis for an experiment by Matthews et al. (in prep.) which in turn is a starting point for the current study. There were two main findings in Matthews et al.'s (in prep.) study that we would like to elaborate on here. Firstly, by modelling the evolution of categorisation using the ILM with human participants (cf. Kirby et al. 2008), they demonstrated how category boundaries are gradually formed during learning in transmission chains. Starting with random category boundaries in a continuous meaning space, learners in transmission chains gradually reorganised them into structured categories. This was done with aid of language labels whose role was to (1) assemble physically similar shapes into similar categories by means of labelling them with similar or identical names, and to (2) separate physically different shapes into different categories by means of labelling them with different labels.

Their second, arguably more exciting, finding was that some participants seemed to be using two different metrics of similarity during language learning. According to one of the metrics, rotation[11] was a relevant factor in categorisation, whereas according to the other it was not. In other words, according to the metric in which rotation mattered, a rotated and an unrotated shape would be considered very different and thus would be given different labels. By contrast, according to the metric where rotation was irrelevant, each of these shapes would be considered similar or identical and thus would be labelled similarly or identically.

Given Matthews et al.'s (in prep.) finding that categorisation systems that evolved in final generations in all four transmission chains had qualitatively different category boundaries, we predict that perception of the meaning space also differed across languages in which these categorisation systems evolved.

This prediction is supported by an abundance of findings demonstrating that cross-linguistic differences in labels used in colour, shape, and other domains can result in cross-linguistic differences in perception of stimuli within and across category boundaries that these labels demarcate (cf. Chapter 2). Note that comparing a number of qualitatively different languages evolved in Matthews et al.'s study with respect to perception of stimuli in the same meaning space is analogue to comparing a number of different real natural languages with respect to perception of a meaning space that is identical for all of them – for example – a meaning space in the domain of colour (cf. §2.3.1). Furthermore, such investigations are akin to testing linguistic relativity suggesting that our perception is influenced by the language we speak (cf. §2.3 on a debate over linguistic relativity).

[11] 'Rotation' is used here, following Matthews et al.'s terminology as shorthand for 'rotation and reflection'.

As discussed in Chapter 2, categorical perception (CP) is the mechanism that allows us to represent our continuous environment as discrete 'chunks' – categories. This is possible because CP shapes representations of categories in such a way that it diminishes within-category differences and increases between-category differences.

There is strong evidence that CP effects (i.e. within-category compression and between-category expansion) are very flexible and undergo developmental changes (cf. §2.4), as well as that they can be artificially induced during categorisation learning in experiments (see §2.5). Given this, we assume that, in Matthews et al.'s study, learning category labels from the output of the previous generation induced CP effects with respect to learnt category boundaries in the learning generation. Furthermore, we assume that the CP effects induced during language transmission were different across languages that evolved in Matthews et al.'s study.

In light of these assumptions as well as the evidence reviewed in Chapter 2, measuring CP effects seems an ideal way of tapping into differences in perception of the meaning space across the evolved languages (i.e., categorisation systems).

There is one apparent problem with respect to testing the prediction that learning different labels that evolved in the categorisation systems in Matthews et al.'s study resulted in differences in perception of categories in the meaning space that these labels referred to. The issue is that Matthews et al.'s experiment was not designed to collect data about perception.

We addressed this problem by taking new participants who learned two of the languages that evolved in Matthews et al.'s experiment and assume that the new participants replace given generations of learners in Matthews et al.'s study. This assumption will allow us to maintain continuity with given transmission chains in their experiment. It is important, however, to replicate the same learning procedures as in Matthews et al.'s experiment so that our replacing generations learn their languages in exactly the same way as their counterparts in Matthews et al.'s study. Once this is done, our prediction can be tested by getting the new replacement generations to do a similarity judgement task on pairs of shapes from the meaning space used in Matthews et al.'s study.

Due to time limits, only two languages evolved in Matthews et al.'s experiment were replaced (cf. *Stimuli for Language Training and Testing Regimen* for a choice justification):

- L1 that replaced the language evolved in Chain 3 in Generation 10
- L2 that replaced the language that evolved in Chain 4 in Generation 10

With this choice in mind, we can state our first hypothesis:

Hypothesis 1

We hypothesise that two shapes in stimuli pairs that

- have been labelled identically in L1 (i.e., categorised into one category) will be perceived as more similar than two shapes in the same stimuli pairs that have been labelled differently in L2 (i.e., categorised into two categories); and that shapes in stimuli pairs that
- have been labelled differently in L1 will be perceived more dissimilar than two shapes in the same stimuli pairs that have been labelled identically in L2.

In other words, we hypothesise that learning two different category systems as represented by L1 and L2 will induce opposite CP effects with respect to the same stimuli pairs. Due to within-category compression and between-category expansion, pairs of shapes labelled identically in L1 will be perceived as more similar than the same pairs of shapes labelled differently in L2, and vice versa.

Our second hypothesis concerns the rotation aspect of languages evolved by Matthews et al. (in prep.). We predict that two metrics of similarity – that is – one that considers rotation relevant and the other that does not – will be very influential with respect to perception of the meaning space, specifically, rotated and unrotated shapes.

Furthermore, we predict that, after learning L1 and L2, the influence of each of the two metrics on perception of the meaning space will be such that it will be possible to capture it in a task where stimuli are not organised according to specific category boundaries as dictated by language labels present in learning data.

The concrete formulation of Hypothesis 2 is as follows:

Hypothesis 2

> Participants learning L1 for which rotation is relevant (i.e., in L1, most rotated and unrotated shapes are named differently) will find rotated shapes less similar than participants learning L2 for which rotation is irrelevant (i.e., in L2, most rotated and unrotated shapes are named identically or similarly).

Apart from evidence supporting linguistic relativity and demonstrating the influence of linguistic development and artificial category learning on perception and CP (cf. Chapter 2 for a review), our experimental investigation has one more, perhaps the most exciting, motivation. L1 and L2 are the languages that evolved during language evolution by cultural transmission. If we are able to observe a significant difference between perceived similarities of items in the meaning space as a result of learning these two languages, then we can support an idea that two qualitatively different semantic systems can evolve during language transmission. Although this finding would be in line with other findings demonstrating co-evolution of language structure with semantics (e.g., Kirby et al. 2008; Matthews et al. in prep.), the current study would be to our knowledge the first to demonstrate the evolution of semantics by means of directly measuring changes in perception. Furthermore, our finding would also be the first to demonstrate that language evolution can introduce different metrics of similarity into the human conceptual system.

5.2 Methods

5.2.1 *Brief description*

27 English-native speakers were recruited from the University of Edinburgh to participate in a study where they had to learn an 'alien language' and perform a similarity judgement task. Participants had no linguistic background at a higher-education level and had never taken part in an 'alien language' learning experiment before. The age mean was 21.56 years; the minimum age was 17 and the maximum 28. The female to male ratio was 17:10. The experiment was designed using E-prime software and it was performed in sound absorbent booths equipped with a PC.

5.2.2 *Stimuli for the Language Training and Testing Regimen*

The stimuli in language training and testing regimen were two languages, consisting of 20 meaning-signal pairs, that evolved in final generations (i.e., Generation 10) in two different transmission chains (Chain 3 and Chain 4) in the study by Matthews et al. (in prep.). We refer to these languages as L1 (i.e., Chain 3 Generation 10) and L2 (i.e., Chain 4 Generation 10). In what follows motivations for the choice of L1 and L2 are supplied, as well as all the relevant technical details with respect to them.

Continuous-meaning space

In the study by Matthews et al. (in prep.) the meanings of all languages, including L1 and L2, come from a continuous 100-element two-dimensional meaning space (Fig.1). Matthews et al. used this kind of meaning space, rather than discrete meanings (e.g., Kirby et al., 2008; cf. § 4.5.1), in order to more accurately model the environment that a language learner categorises with aid of language labels.

In the meaning space each image varies in height, length, shape and rotation. In two opposite corners of the meaning space, there are a rectangle and a rectangle that is a 90° rotation of it, whereas in the other two there are an equilateral triangle and a triangle that is a 90° rotation of it. The remaining images in the meaning space were morphed from the images in each corner, depending on distance from the corners (cf. Matthews et al. in prep. for further details.).

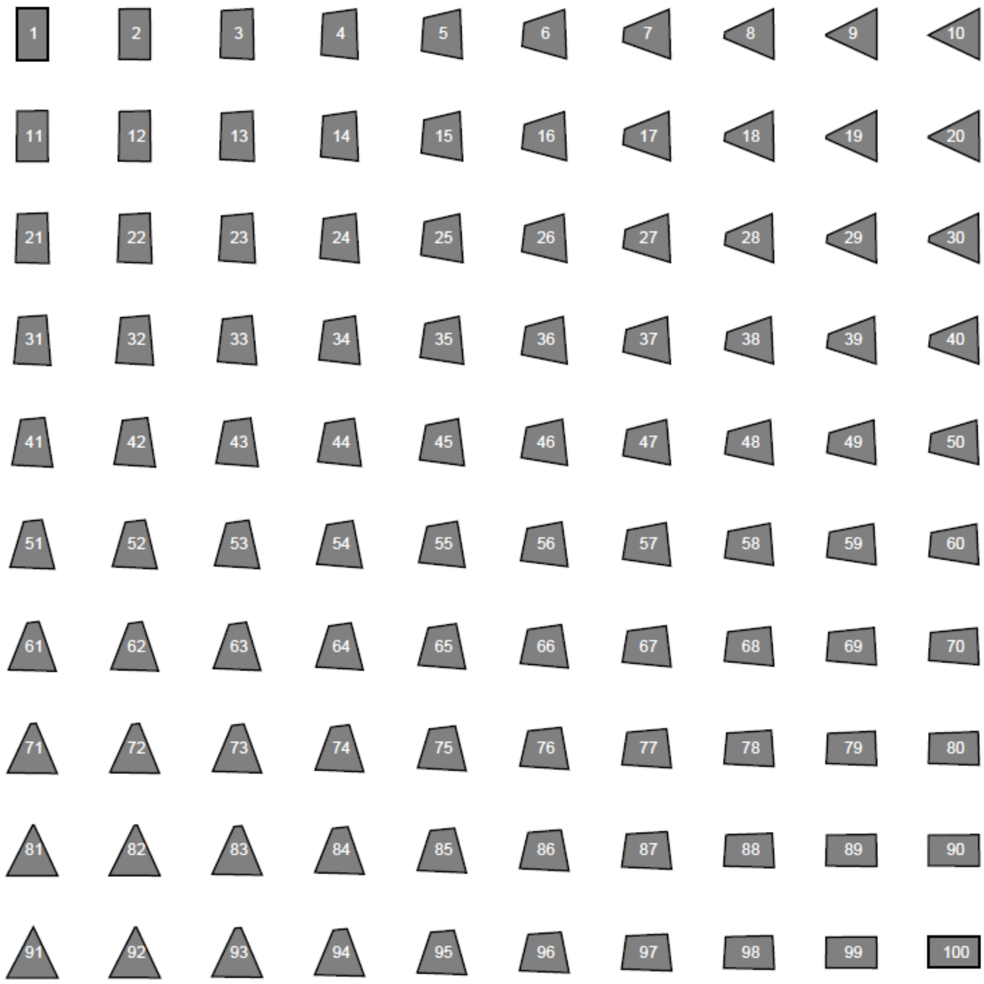


Fig.1 The full 100-element meaning space from which meanings in L1 and L2 come. Reprinted with permission from Matthews et al. (in prep.).

Transmission chains

There were 4 transmission chains in Matthews et al.’s study, each consisting of 10 generations of learners (each generation was represented by a single participant). All generations learned a language consisting of 20 signal-meaning pairs. The first generation in each chain learned a random language, consisting of 20 signals (strings between 2 and 4 syllables that were randomly concatenated out of 9 syllables) that were randomly assigned to 20 meanings, randomly chosen from the 100-element continuous meaning space. All remaining generations learned a language that was the previous participant’s output in the final test.

Motivations of the choice of L1 and L2

Experimental procedure in Matthews et al.’s (in prep.) experiment consisted of three rounds of training (cf. *Language Training and Testing Regimen* below for a detailed description) and a final test. The final test comprised 10 Seen items (the meanings that participants had already seen during training) and 10 Unseen items (the meanings that participants had not seen during training), and the set of 16 fixed items. The set of 16 fixed items consisted of shapes that were evenly distributed in the meaning space (see Fig. 2).The set was the same for all participants, thus constituting a good sample for obtaining relevant measurements that could be compared across generations in Matthews et al.'s study.

The stimuli used in the current experiment (i.e., meaning-signal pairs comprising L1 and L2) were obtained from the final test in Generation 10, in both Chain 3 and Chain 4 in Matthews et al.'s experiment. However, L1 and L2 only comprised 10 Seen and 10 Unseen items, without the fixed set.

The choice of 20 items from the final test, instead of, for example, 20 items from earlier tests during 3 rounds of training, is motivated by the fact that the former were the language stimuli serving as the input for the next learning generation in Matthews et al.'s study. Furthermore, the 20 items from the final test are the most advanced form of an evolved language in Matthews et al.'s experiment because they reflect how some language items were memorised (Seen items) and how some were generalised and/or innovated (Unseen items) after the same amount of exposure to training data. Therefore, the 20 items from the final test could be claimed to model the result of learning processes usually involved in language evolution in the ILM.

Using categorisations of the set of 16 fixed items, Matthews et al. (in prep.) calculated language structure for each generation in all 4 chains.

A language is structured if shapes which are close to each other in the meaning space have similar (or identical) names and shapes which are far away from each other have different names. In order to establish name similarity and shape distance Matthews et al. used two measures: the Levenstein Distance (LD)[12] and the Euclidean Distance (ED), respectively. If these two measures, calculated for the set of 16 items, are positively correlated, then the language can be thought of as structured. A distribution of correlation values for Monte-Carlo permutations of signal-meaning pairs of the fixed set was used to obtain the statistical significance of the language structure measurement. The approach to structure measurement used by Matthews et al. is known as the Mantel test.

The choice of L1 and L2 for the current study was based on the significance of structure measures calculated in Matthews et al.'s study. This was motivated by the fact that significant measures of structure guarantee that L1 and L2 are substantially clustered (i.e., structured), thus maximising our chances of obtaining meaningful data in the similarity judgement task.

Apart from Matthews et al.'s finding that evolved languages became gradually more structured by means of categorisation, the authors also found that rotation was a relevant dimension for categorisations in final generations. They found that in some languages (R-languages) two shapes

[12] Levenstein distance (or edit distance) (Levenstein, 1966) is calculated as the number of edit operations (insertions, deletions, or substitutions) needed to be performed to transform one string to another.

which were rotated and reflected were named identically (i.e., categorised into one category) or very similarly (i.e., categorised into similar categories), whereas in other languages (NR-languages) they were named differently (i.e., categorised into two different categories) (cf. Fig. 3).

In order to take into account the differences with respect to rotation when calculating language structure, Matthews et al. used two different measures of shape distance – that is – two different EDs. Given this, they obtained two different measures of structure:

- Non-rotational (NR) Structure, for which rotated and reflected shapes are named differently, was calculated using ED1:

$$ED1(p,q) = \sqrt{(x_2-x_1)^2+(y_2-y_1)^2}$$

- Rotational (R) Structure, for which they are named the same or very similarly, was calculated using ED2:

$$ED2(p,q) = min\left(\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}, \sqrt{(x_1-(9-x_2))^2+(y_1-(9-y_2))^2}\right)$$

Our second hypothesis concerns testing whether perception of rotated and reflected shapes is different in participants who learnt two languages that use different similarity metrics regarding rotation differently. Therefore, in our search of L1 and L2 we are determined to find two languages that differ the most with respect to how they categorise rotational shapes. This implies that L1 should have the highest significance of NR-structure, and a non-significant R-structure, whereas L2 should have the highest significance of R-structure, and a non-significant NR-structure. Accordingly, the choice was made for L1 to have NR-structure[13] = 4.43, R-structure = 0.01; and for L2 to have R-structure = 6.18, NR-structure = 0.97.

[13] Measure of significance of structure – cf. Matthews et al. (in prep.) for details.

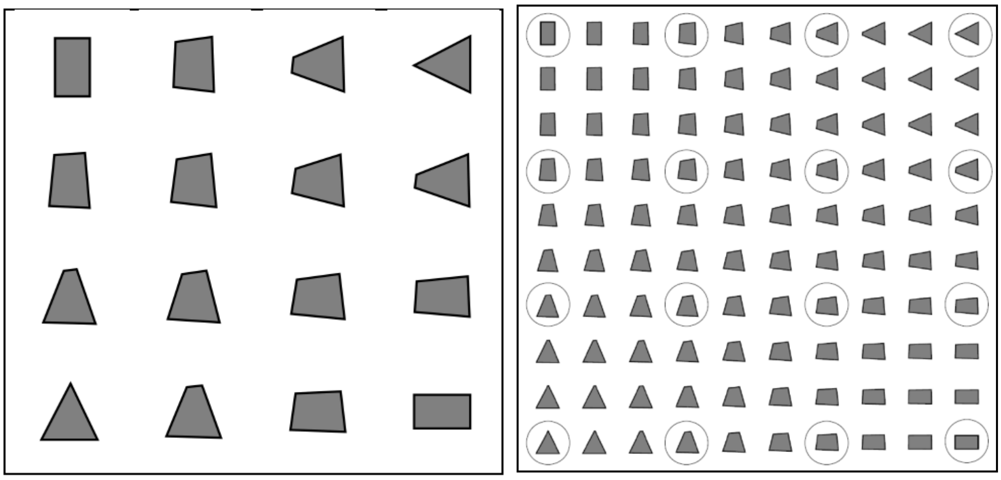


Fig.2 The set of 16 fixed items (left). The position of fixed items in the meaning space marked with a circle. Reprinted with permission from Matthews et al. (in prep.).

13 88

Fig.3 An example of a shape pair containing an unrotated (left) and rotated shape (right). In R-languages such shapes were named identically or similarly, whereas in NR-languages they were named differently. Reprinted with permission from Matthews et al. (in prep.)

5.2.3 *Stimuli in the Similarity Judgement Task*

The stimuli in the similarity judgement task consisted of 80 pairs of shapes from the 100-element meaning space used in Matthews et al.'s experiment (see Appendix C for the full list of stimuli). Although there were 4950 possible combinations of shapes in a pair, we limited the number of stimuli pairs that we used to 80. This limit was enforced to ensure that participants remained focused for the duration of the experiment, thus preventing random decisions about shape similarity.

The 80 pairs can be divided into 2 sets of data (56-pair and 24-pair sets) based on how they were analysed. The 56-pair set was used in the data analysis concerning Hypothesis 1, whereas the 24-pair set was used in the analysis concerning Hypothesis 2.

Stimuli for testing Hypothesis 1

The set of 56 shape pairs consisted of 28 horizontal and vertical and 28 diagonal pairs (cf. Fig. 4). In order to be chosen from the meaning space, all shape pairs in this set had to meet the following criteria:

- ED1 between shapes in each pair should be equal to ED2 (criterion 1)
- Both ED1 and ED2 should be approximately equal for all types of pairs (i.e. horizontal, vertical and diagonal) (criterion 2)
- Stimuli pairs should consist of two types of pairs:
  - pairs in which both shapes should be labelled identically in L1 and differently in L2 (i.e., both shapes in a pair should belong to the Same category in L1, henceforth S pairs, and to Different categories in L2, henceforth D pairs); we refer to such stimuli pairs as SD pairs;
  - pairs in which both shapes should be labelled differently in L1 and identically in L2; we refer to such stimuli pairs as DS (criterion 3)
- The number of SD pairs and DS pairs should be the same; this means that the number of S pairs and D pairs in each language (i.e., L1 and L2) should be the same (criterion 4)

The Euclidean distances, ED1 and ED2, were 3 for horizontal and vertical pairs and 2.8 for diagonal pairs which makes both types of distances approximately equal across all three types of pairs.

Although we realise that equal Euclidean distances do not correspond to equal psychological distances (i.e. equal discriminability), we took into account Criteria 1 and 2 in order to attempt to lessen variability with respect to psychological similarity between the stimuli. However, the maintenance of equal Euclidean distances within pairs of stimuli was not strictly necessary because label learning effects on perceived similarity of shape pairs in L1 were assessed by means of comparing them to parallel effects on perceived similarity of exactly the same pairs in L2. In other words, discriminability differences between pairs of stimuli should not confound our measurements as long as they influence perceived similarity in both L1 and L2 equally.

Our aim is to examine the hypothesis that perceived similarity of two shapes within a pair that have been categorised into one category in L1 is larger than perceived similarity of the same shapes that have been categorised into two categories in L2 and vice versa.

In order to test this hypothesis, we required that each shape in each stimuli pair be categorised according to criteria 3 and 4. Ideally this categorisation (i.e., labels for each shape) would come from learners after they have trained on L1 and L2. This suggests that stimuli for the similarity judgement task should come from the final test on the set of 16 fixed items performed by each participant in our study (cf. *Motivation of the choice of L1 and L2* for more details of what the set of 16 fixed items was;

and *Language Training and Testing Regimen* for details of experimental procedures). However, cross-participant variation with respect to categorisation of the same items in the set would make it extremely difficult to deduce a uniform way of categorisation of the fixed set.

The problem of cross-participant variation was resolved by means of using categories that a perfect learner would generate after learning L1 and L2[14] (see Appendix A for the full list of categorisations made by a perfect learner). A perfect learner simulated here remembers every signal-meaning pair it has seen and then labels a novel meaning (i.e., a shape) with the name given to the nearest meaning it has learnt. This type of learner is an example of a 'k-nearest neighbour classifier' with k=1, where 'k' is the number of nearest seen items that are considered when deciding how to categorise novel meanings.

Using categorisations of a perfect learner we are no longer confined by the necessity of using the set of 16 fixed items as the source of categories. Instead, 100 labels were generated for each shape in the 100-element meaning space by simulating a perfect learner. This allowed us to choose the 56 pairs of stimuli from the 100-element meaning space, keeping criteria 1-4 in mind.

The issue of the equal number of SD pairs and DS pairs (criterion 4), was resolved in the following way. Firstly, all horizontal and vertical pairs with ED1=ED2=3 and diagonal pairs with ED1=ED2=2.8 were collected from the meaning space. Then the number of SD and DS pairs was calculated for the horizontal and vertical group and for the diagonal group. The number of DS pairs in the horizontal and vertical group was 14. However, because the number of SD pairs in the horizontal and vertical group and the number of SD and DS pairs in the diagonal group exceeded 14 we randomly chose 14 pairs of each relevant pair type. As a result there were 14 SD and 14 DS pairs in the horizontal and vertical group and 14 SD and 14 DS pairs in the diagonal group.

[14] Simon Kirby has kindly provided categories generated by a perfect learner for L1 and L2.

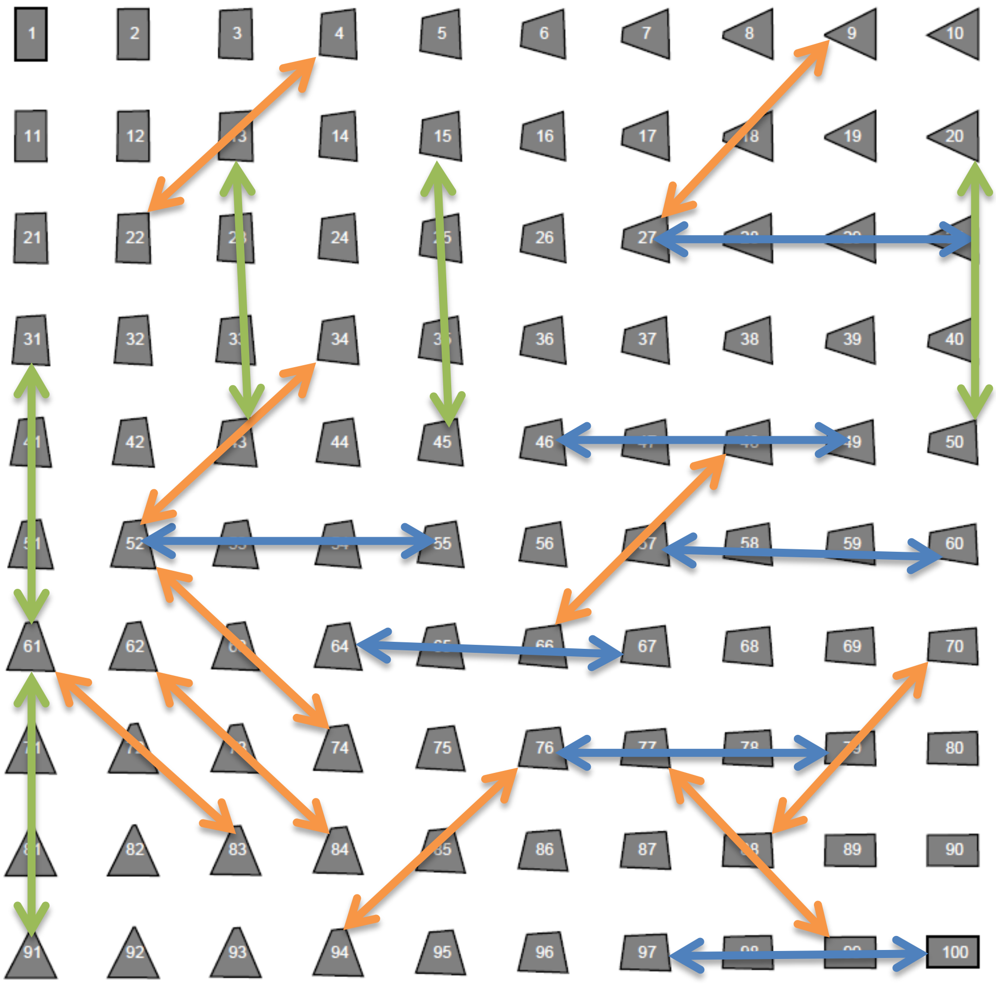


Fig.4 Examples of stimuli pairs used in the similarity judgement task to test Hypothesis 1. Horizontal, vertical and diagonal pairs are indicated by blue, green, and orange arrows respectively.

Stimuli for testing Hypothesis 2

The set of 24 shape pairs exclusively consisted of pairs in which one shape was a rotation of the other (cf. Fig.5). In other words, we chose pairs of shapes for which ED1 was as high as possible and ED2=0.

This choice of the 24 pair shapes enables us to focus on the perceived similarity of pairs of shapes that differ solely along the dimension of rotation, thus comprising an ideal set for testing Hypothesis 2.

It is also worth noting that for the purpose of testing Hypothesis 2, we did not have to organise the stimuli in each pair into categories as we did with the stimuli for Hypothesis 1. This is because we expected that we would observe the influence of both similarity metrics (the one that causes participants to consider rotation relevant and the one that does not) regardless of how categorical boundaries of testing stimuli were organised.

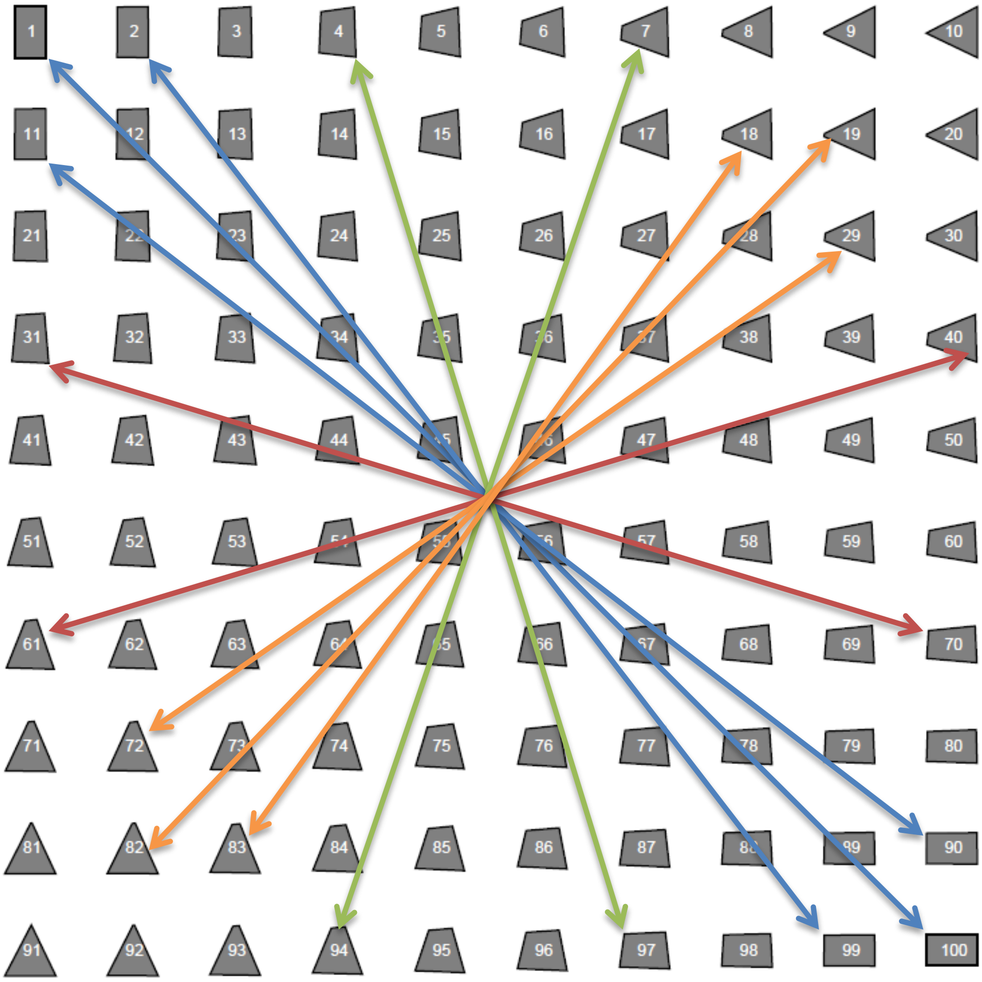


Fig.5 Examples of stimuli pairs used in the similarity judgement task to test Hypothesis 2.

5.2.4 *Procedure*

General outline

The experimental procedure consisted of three rounds of language training, with two minute breaks between each round, a similarity judgement task, and a final language test (cf. Appendix B for experimental instructions). Apart from the similarity judgement task, all parts of the experimental procedure were adapted from Matthews et al.'s (in prep) experimental procedure[15].

Language Training and Testing Regimen

[15] The code for language learning and training tasks was kindly provided by Cristina Matthews. The code for the similarity judgement task was prepared by the author.

Participants were randomly divided into two groups – one trained on L1 (14 participants) and another trained on L2 (13 participants). In the first two training rounds participants were exposed to 20 randomly ordered signal-meaning pairs (comprising all the L1 or L2, depending on the condition) and were tested on 10 out of 20 randomly chosen meanings for which they had to supply signals. In the third training round the training procedure was the same, but there was no testing. After all the three rounds of training participants performed the similarity judgement task and the final test on the set of 16 fixed items from the study by Matthew's et al. (in prep.).

The role of the final test was to produce a metric for measuring accuracy of all participants (cf. §5.3.3).

Similarity Judgement Task

Participants in both groups (L1 and L2) judged similarity of 80 pairs of shapes that were chosen according to the criteria explained in §5.2.3. After the final round of language training participants were instructed on the computer screen that they would be asked to judge the similarity of pairs of shapes. During each trial a pair of shapes appeared on the computer screen. Both shapes in each pair were positioned side by side and appeared on the computer screen simultaneously for 2,500 ms. The simultaneous, as opposed to consecutive, appearance of both shapes was chosen because it has previously succeeded at being a reliable method demonstrating CP effects in normal and language-impaired participants (e.g., Roberson, Davidoff, & Braisby, 1999). Furthermore, if two shapes appear simultaneously, short-term memory is burdened less than when they appear consecutively.

After each pair had disappeared, participants marked, on the similarity judgement sheet provided, the similarity of the pair they had just seen. In each similarity judgement sheet there were 5 lines per page, comprising together 80 lines. Each line was 12 centimetres long with the beginning and end points marked as 'not similar' and 'very similar' respectively. Such a line enables continuous measurements. All participants judged the similarity of each of 80 pairs and the order of appearance of all the pairs was randomised and different for each participant.

After completion of the experiment, each similarity judgement was measured by the experimenter and rescaled such that the markings at the end points were 1 ('not similar') and 7 ('very similar'). The rescaling was done in order to obtain measurements with numbers of magnitude parallel to that used in 7-point Likert scales.

5.3 Results

5.3.1 *Hypothesis 1*

A paired-samples t-test was conducted in order to compare similarity ratings of shape pairs where both shapes in a pair were labelled identically in L1 and L2 (Same condition) with similarity ratings of matched shape pairs where both shapes in a pair were labelled differently in L1 and L2 (Different condition). There was no significant difference between Same (M=3.797, SD=0.86) and Different (M=3.810, SD=0.9) conditions; $t(55)=-0.28$, $p=0.78$.

This result demonstrates that there was no difference between perceived similarity of shapes in pairs where both shapes were in one category in L1 and in two categories in L2, as well as of shapes in pairs where both shapes were in two categories in L1 and one category in L2. This further means that our hypothesis that similarity ratings in Same condition would be higher than in Different condition is unsupported.

5.3.2 *Hypothesis 2*

Another paired-samples t-test was performed in order to compare, between conditions L1 and L2, similarity ratings of 24 shape pairs, where one shape in a pair was a rotation of the other. It was hypothesised that similarity ratings will be higher in condition L2 than in condition L1 because participants learning L2 were exposed to a considerable number of signal-meaning pairs for which rotation was irrelevant (i.e., two shapes varying in rotation were labelled similarly or identically), as opposed to participants learning L1 who were exposed to a considerable amount of signal-meaning pairs for which rotation was relevant (i.e., two shapes varying in rotation were labelled differently). We found a significant difference between condition L1 (M=5.13, SD=0.4) and condition L2 (M=5.61, SD=0.6); $t(23)=-3.41$, $p=0.002$.

This finding suggests that participants learning L1, for which rotation was relevant, found shape pairs, in which one shape was a rotation of the other, perceptually different than participants learning L2, for which rotation was irrelevant. Furthermore, our hypothesis that similarity ratings in L1 would be significantly lower than in L2 has been supported (M of L1 < M of L2).

5.3.3 *Post-hoc analysis*

A post-hoc analysis was performed with respect to testing Hypothesis 1 in order to remove possible outliers. We assumed that similarity ratings could be influenced by how well participants had learnt the languages. In order to investigate this we measured participants' accuracy in categorisation after learning. Given this, outliers were defined as those learners who did not categorise the set of 16 fixed items in the final test sufficiently accurately. Accuracy scores were established by means of calculating the number of participants' labels (categories) given to shapes in the fixed set that were identical to the labels given to the same shapes by the perfect learner.

9 participants in each group (i.e., L1 and L2) with the highest accuracy scores were chosen for a post-hoc statistical analysis. In other words, participants with scores lower than 8/16 in L1 and 4/16 in L2 were excluded as outliers.

A paired-samples t-test was performed as the post-hoc statistical analysis to compare similarity ratings between Same and Different conditions (cf. the § on the results for Hypothesis 1 for more details on the conditions). There was no significant difference between the conditions (Same: $M=3.93$, $SD=0.83$; Different: $M=3.89$, $SD=0.86$); $t(55)=0.57$, $p=0.57$.

The result of the post-hoc statistical analysis demonstrates that even after the removal of outliers there was no difference between perceived similarity between Same and Different conditions.

Chapter 6 Discussion

6.1 Hypothesis 1

6.1.1 *Design issues*

We hypothesised that perceived similarity of shapes in pairs that were categorised identically in L1 and L2 (Same condition) would be higher than perceived similarity of matched shapes in pairs that were categorised differently in these languages (Different condition). Our hypothesis was motivated by findings supporting the Whorf hypothesis that speaking languages which differ in how they categorise the meaning space results in differences in perception of this space (cf. Chapter 2 for a review). Therefore, we expected that learning L1 and L2, which differ in how they categorise the meaning space, would result in within-category expansion and between-category compression of perceived similarity. It is worth noting that it is a common practice in psychological literature (cf. Chapter 2) to test linguistic relativity by means of measuring CP effects.

Our result failed to demonstrate a difference in perceived similarity between Same and Different conditions. This could be interpreted as indicating that the effect of language learning (i.e., training) was not strong enough to influence and change perceived similarity of the stimuli. In other words, it is possible that there was too little language training in the current experiment. Such an interpretation is motivated by an observation that in other studies, that successfully demonstrate changes in perceived similarity of stimuli as a result of assigning categories to the stimuli, there is substantially more training.

For example, in Goldstone's (1994a; Experiment 2) study demonstrating that perception of size and brightness changes as a result of assigning two categories to stimuli characterised by these dimensions, participants are exposed to 20 repetitions of all 16 stimuli (i.e., object-category pairs) during training. In another study investigating how categorisation learning changes perceived similarities between faces (Goldstone et al., 2001) each stimulus was presented 54 times. By contrast, training in our experiment comprised only 3 learning rounds, hence each training stimulus was presented only 3 times.

Another factor that could have contributed to the finding that no reliable difference was present in perceived similarity between Same and Different conditions could be the fact that no feedback was supplied during training in the current experiment. Although unsupervised learning (i.e., without feedback) is involved when we learn our native language, supervised learning (i.e., with feedback) is typically more successful when it comes to demonstrating CP effects as a result of artificial category learning (e.g., Goldstone, 1994a; Goldstone et al., 2001). Therefore, supplying feedback to participants during language training could have potentially resulted in better learning of L1 and L2, and, as a consequence, it could improve our result.

Another problem that could have contributed to the non-significant result we obtained concerns the fact that, in a typical study designed to investigate changes in perceived similarity as a result of category learning, participants are trained on exactly the same number of stimuli that are subsequently used in the similarity judgement task (e.g., Goldstone, 1994a; Goldstone et al., 2001). In the current experiment this was not the case.

We assumed that training on only a subset of the stimuli subsequently used in the similarity judgement task (i.e., 13/112 for L1 and 14/112 for L2[16]) would be enough to influence participants' perception of the inexperienced elements in the meaning space. Furthermore, we assumed that, because the stimuli used in the similarity judgement task lay within category boundaries that were to be identified during training, it would be possible for participants in the similarity judgement task to generalise over similarities of shapes they did not see during category learning.

However, it is possible that the subset of the stimuli used in the similarity judgement task, that participants were trained on, could contain too few category exemplars. The implication of this possibility is that, with such a limited number of training exemplars, participants could develop category representations that were too weak to enable them to induce similarity of shape pairs they had not experienced during training.

It is worth emphasising that all the problems mentioned so far (i.e., too little training, no feedback during training and the lack of identity of stimuli used during training and the similarity judgement task) are a consequence of a trade-off between a desire to maintain continuity with Matthews et al.'s (in prep.) experiment and a desire to design an experiment that would investigate the influence of learning two different languages on perception. Maintaining continuity was crucial since we set out to investigate the idea that language evolution, as modelled in their study, can lead to the emergence of two qualitatively different semantic systems.

Furthermore, we need to bear in mind that, although the limited training could have indeed contributed to the non-significant result in the current experiment, the training was sufficient in Matthew's et al. experiment to result in the evolution of categorisation. The reason for this could be that the alteration of perception may take place at a more deeply embedded level of processing, thus requiring more training to be effective. On the other hand, the evolution of categorisation without any changes in perception could take place at a level that requires less training.

[16] Because L1 and L2 are from Matthews et al.'s study, we did not have any influence on neither the number of signal-meaning pairs in them (there were 20), nor what they were. Thus, the fact that only 13 items from L1 and 14 from L2 were present in the set of stimuli for testing Hypothesis 1 in the similarity judgement task was beyond our control.

6.1.2 *A perfect learner problem*

The choice of the perfect learner as the source of category labels used to generate Same/Different stimuli to test Hypothesis 1 (cf. *Stimuli for Testing Hypothesis 1*) may have also influenced our result.

The perfect learner helped us to eliminate the problem of cross-participant variation (cf. *Stimuli for Testing Hypothesis 1*). Nevertheless, the results of the accuracy test conducted for the post-hoc analysis demonstrate that participants' performance in labelling categories in the set of 16 fixed items, measured relative to the performance of the perfect learner, was not very good, with the highest scores of 12/16 for L1 and 8/16 for L2 and the lowest scores of 3/16 for L1 and 2/16 in L2 (cf. Fig. 6a-b)[17].

Such a low level of accuracy could indicate that participants did not learn the languages very well and as a consequence they did not form category representations strong enough to drive differences in perceived similarity of stimuli between L1 and L2.

However, such a low level of accuracy could also suggest that the perfect learner's performance was too rigid a reference point for participants' performance because human learners are far from perfect learners. Therefore, it could be possible that the categories to test Hypothesis 1 might have been wrongly chosen. This could have resulted in a mismatch between categories that were formed in participants' heads during training and the categories obtained from the perfect learner that they were tested on. As a consequence, we obtained a non-significant result and, possibly wrongly, interpreted it as indicative of weak learning performance.

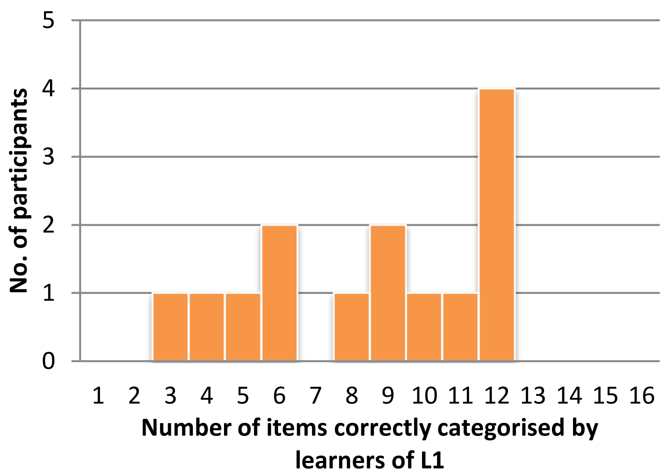


Figure 6a: Accuracy of categorisation of the set of 16 fixed items by participants learning L1 measured by the number of items that a participant labelled identically as the perfect learner.

[17]For simplicity, we adopted a simple accuracy test which does not take into account differences in Levenstein distances between the labels produced by participants and a perfect learner.

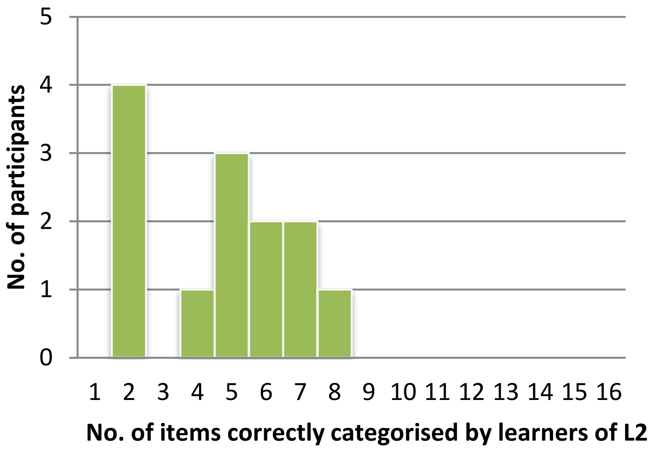


Figure 6b: Accuracy of categorisation of the set of 16 fixed items by participants learning L2 measured by the number of items that a participant labelled identically as the perfect learner.

6.2 Hypothesis 2

Our second hypothesis that learning languages that use two different metrics of similarity, one, for which rotation is relevant (L1) and another, for which it is not (L2), would result in differential perception of rotated and unrotated stimuli has been found to be supported by the experimental data. Furthermore, the perceived similarity of stimuli pairs in which one shape is a rotation of the other was found to be higher in L2 than in L1, as we predicted.

Given our result, it seems clear that the difference in perception of rotation between L1 and L2 was caused by learning stimuli that consisted of rotated and unrotated shapes labelled differently in L1 and similarly or identically in L2. Furthermore, we can conclude that our result is robust because the difference in perception of rotation between the languages was observed despite different category boundaries in each language (i.e., despite Different/Same categorisations made by the perfect learner, as discussed earlier).

Upon a closer examination of training stimuli in L1, we can observe that, although most stimuli consist of rotated and unrotated shapes that are labelled differently (e.g., orange and light green shapes and yellow and light green shapes in Fig.7), there are some rotated and unrotated shapes (i.e., dark green shapes in Fig.7) that are contradicting the language's general tendency because they are labelled identically. However, our result demonstrates that despite these counterexamples in L1 training data, the influence of L1 on perception of rotation is still detectable. By contrast, in L2 there are no counterexamples to the language's tendency to name rotated and unrotated shapes identically or similarly (e.g., cf. Fig.8 where all yellow, orange and red shapes are named similarly and all black, dark, light and very light blue shapes are named similarly).

In the context of our results that differ with respect to Hypothesis 1 and 2, it seems relevant to attempt to answer the question of why Hypothesis 2 has been supported, whereas Hypothesis 1 has not.

Firstly, it is interesting to note that the influence of L1 and L2 on perception of rotated and unrotated shapes was so robust that it was uncovered despite all the issues that constrain the possibility of finding support for Hypothesis 1. In other words, such problematic issues as limited training, no feedback during training and the lack of identity of stimuli used during training and in the similarity judgement task do not seem to play a role with respect to testing Hypothesis 2 at all (See *Design Issues* for a discussion of these issues with respect to Hypothesis 1). Why is it the case that the influence of L1 and L2 on perception of rotation is so strong, whereas the influence of L1 and L2 on perception of within-category and between-category stimuli is not?

The answer to this question we suggest is related to the nature of mechanisms involved in category learning and to the fact that learning in the experiment was restricted. It seems that in such a case there needs to be a perceptually salient feature in training stimuli that would facilitate category learning. This would imply that such a feature could be quickly picked up by perceptual learning mechanisms and easily emphasised or de-emphasised by association with language labels during training.

Category learning involves mechanisms whereby relevant perceptual features are emphasised or de-emphasised by respective assignment of identical or different labels to stimuli containing them. In circumstances where training is long (cf. § 6.1.1 for examples), category learning probably does not need to be aided by presence of a salient perceptual dimension because there is enough training to enable formation of more complex statistical relations between perceptual features of training stimuli and labels assigned to them. However, in a situation where training is limited, as in our experiment, only a feature that is perceptually salient can be picked up by language labels during training. In the current experiment rotation is such a salient perceptual feature.

Perceptual salience of rotation is motivated by the fact that orientation, crucial for representation of rotations, is one of the first perceptual dimensions to be processed in the primary visual cortex (Hubel & Wiesel, 1962; Paradiso, 2002)[18].

In light of the above considerations, it is possible that rotation was the only perceptually salient stimuli feature that could be associated with language labels during training. As a consequence, category learning was facilitated and successful, and rotation was recognised as perceptually salient again when Hypothesis 2 was tested in the similarity judgement task.

[18] Hubel and Wiesel (1962) discovered orientation-sensitive cells in the cat visual cortex. Paradiso (2002) reviews results demonstrating correlations of neuronal activity in the visual primary cortex and perception of a number of features, including orientation.

Furthermore, it is possible that there was not enough training to enable participants to recognise any other perceptual features, less distinctive than rotation, that could facilitate category learning. As a result, participants were unable to successfully learn categorisation and gave random similarity ratings to the stimuli for testing Hypothesis 1 in the similarity judgement task.

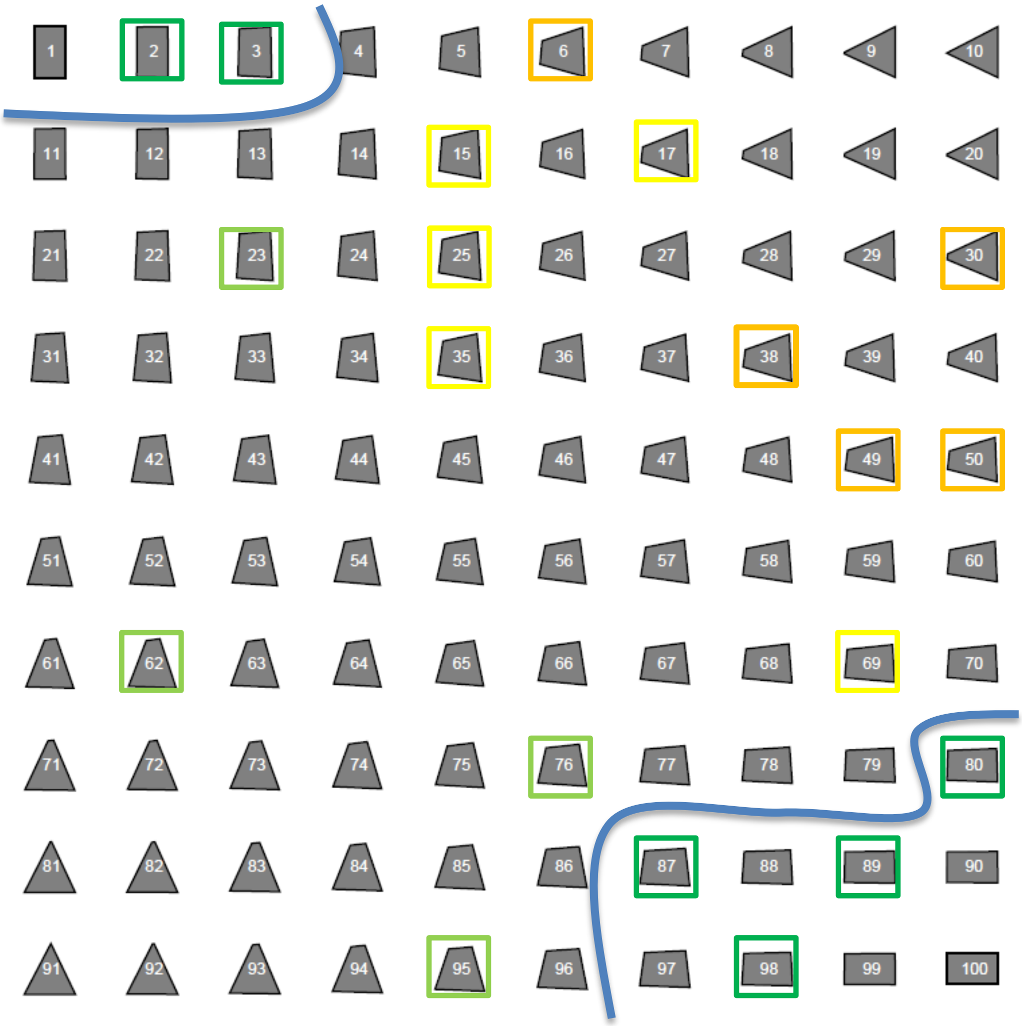


Fig.7 Shapes from L1 training. Coloured squares indicate identical labels (orange - nikihe; light green – pani; yellow – nikihehe; dark green – mani). Blue lines demarcate counterexamples to the language's general tendency to consider rotation relevant (i.e., to name rotated and unrotated shapes differently).

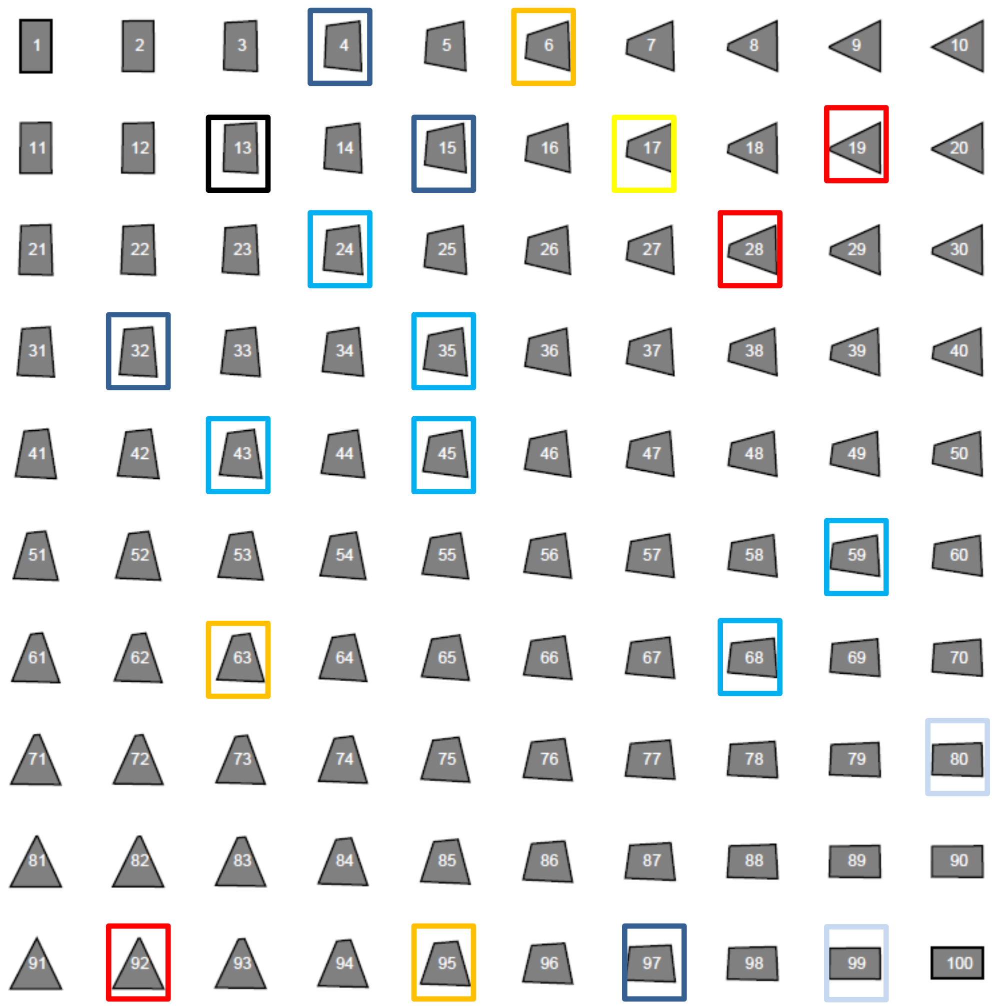


Fig.8 Shapes from L2 training. Coloured squares indicate identical labels (dark blue – mauhihi; light blue – maukihi; very light blue – maukiki; orange – maumauage; yellow – maumage; maumag – red; hihi – black).

6.3 General problems

6.3.1 *Conceptual processing confound*

In the current experiment we were interested in the influence of L1 and L2 learning on perceived similarity of the meaning space. However, similarity ratings used to measure perceived similarity in the current experiment are prone to reflecting conceptual (i.e., high-level), rather than perceptual (i.e., low-level) processing (Goldstone, 1994b; Goldstone et al., 2001). For example, similarity ratings can reflect factors, such as, goals and intentions of the comparison maker and the comparison maker's expertise in dealing with geometrical figures due to her field of study or job (Goldstone, 1994b), as well as the strategic use of language labels (Goldstone et al., 2001). In the current study, the strategic use of category labels seems to be the most problematic on the account of the results of a post-experimental questionnaire.

The questionnaire was conducted immediately after participants completed the experiment. They were enquired about criteria they took into account when rating similarities of shape pairs in the similarity judgement task. More specifically, they were asked whether they considered geometrical properties of shapes in the task at hand or labels of shapes learnt during training when they rated similarity of stimuli. To enable participants to understand what was meant by 'consideration of labels learnt during training', they were given an example of a thought they could have: 'I will give these two shapes a high similarity rating because just a moment ago I labelled them using the same name'.

Although most shapes used in the similarity judgement task were not included in the training part of the experiment, they were within boundaries of categories to be learnt during training. Thus, we assume that due to this closeness to learnt category boundaries, participants could still consciously take into account labels of the learnt stimuli when judging similarity of the unseen stimuli.

The questionnaire results demonstrate that 13/27 participants consciously thought in at least 50% of cases about language labels when judging similarity of shape pairs. On the other hand, 22/27 participants consciously thought about geometrical properties in at least 50% of cases.

Although we did not find any significant difference between L1 and L2 with respect to similarity ratings of Same/Different shape pairs in the current experiment (Hypothesis 1), it seems relevant to bear in mind the strategic use of linguistic category labels as a possible confound in future experiments. Even more importantly, however, our result that L1 and L2 differentially influence perception of rotation (Hypothesis 2) would be stronger, if we had controlled for the strategic use of linguistic labels.

As mentioned earlier, the reason for this is that, ideally, we would like to exclude any explanation of our results that involves high-level processes as accounting for the difference in perceived similarity of stimuli. Instead, we are interested in low-level processes as an explanation. Thus, controlling for the strategic use of linguistic labels would effectively result in removal of one of the most obvious confounds related to high-level processing.

One way to avoid the confound of the strategic use of language labels in the procedure used for testing Hypothesis 1 is to ask for similarity ratings relative to some uncategorised shape (adapted from Goldstone et al., 2001). The idea here is that, given two shapes A and B belonging to the same category in L1 and to different categories in L2 and an uncategorised item E, we predict that, if A and B are judged to be more similar in L1 than in L2 (thus, diminishing distance between them in the similarity space), then, the difference between similarity judgements between the pair of A and E and the pair of B and E will be smaller in L1 compared to L2. A similar way of controlling for the strategic use of linguistic labels would need to be invented for stimuli used in testing Hypothesis 2.

Another way of controlling for the possibility of the involvement of conceptual processing would be to change the type of the similarity judgement task entirely. It has been suggested that the best tasks that allow us to control for conceptual processing are perceptual discriminability tasks (Hodgetts, Hahn, & Chater, 2009). The following are examples of such tasks that could be used in future experiments:

- *an odd-one-out matching triad task*, where two most similar stimuli are chosen out of three; stimuli are physically equidistant (or equally discriminable) and typically their category membership is manipulated to achieve desired CP effects (e.g., Roberson et al., 2000)

- *a two-alternative forced choice task*, where, on a typical trail, the stimuli consist of two pairs of objects and participants are required to say whether similarity between objects in pair 1 is greater than in pair 2 (e.g., Roberson et al., 2000; Hodgetts et al., 2009)

- *a same/different discrimination task* where 2 stimuli are judged as same or different; stimuli are physically equidistant (or equally discriminable) and typically their category membership is manipulated to achieve desired CP effects (e.g., Goldstone, 1994a)

- *an ABX discrimination task* where participants are exposed to stimuli A, B and X in succession and required to say whether X was identical to A or B; X is always identical to A or B; category membership is manipulated for A and B to achieve desired CP effects (e.g., Liberman et al., 1957)

Although our intention here is not to give exhaustive explanations of perceptual discriminability tasks, the above list could serve as starting point for possible future modifications of the current experiment.

6.3.3 *Similarity construct*

When designing the current experiment we assumed that changes in perception can be quantified by means of measuring changes in perceived similarity of stimuli. Therefore, when making experimental design choices we should have been guided by a psychologically plausible notion of similarity.

Given the complexity of this problem, which was partially due to the constraints imposed by the meaning space from Matthews et al.'s experiment and the inherent ED measure of similarity, we designed the similarity judgement task without having any particular, psychologically motivated, notion of similarity in mind.

However, given the non-significant result with respect to Hypothesis 1, it is possible that what was needed was a specific notion of psychological similarity to guide our design of stimuli for the similarity judgement task. This possibility is strengthened by the fact that recent evidence suggests that geometrical (e.g., Shepard, 1957) or featural accounts of similarity (e.g., Tversky, 1977) are not psychologically plausible because they assume that similarity is sensitive to single features of objects and not interactions between them. Although to test Hypothesis 2, a single feature of rotation was the most relevant, it was unclear which features and featural interactions (e.g., the number of sides and pointiness) would be relevant to testing Hypothesis 1.

Instead of assuming geometrical or featural notion of similarity, recent evidence demonstrates that similarity between pairs of geometrical objects is best captured by the number of complex distortions that need to be accomplished in order to change one object representation to another (the Representational Distortion account – Hahn, Chater, & Richardson, (2003)). In light of this evidence, it may be worthwhile to assume the Representational Distortion notion of similarity and design the similarity judgement task stimuli accordingly.

The issue of choosing the most plausible notion of psychological similarity is very complex and hotly debated in psychological research (e.g., Goldstone, 1994; Goldstone, Day, & Son, 2010). Thus, it is not our intention to provide a solution to this problem, but merely to suggest its existence, and possible relevance to the current experiment.

6.4 Significance of the experiment to language evolution

The leading motivation for the current experiment was the idea to conjoin two research fields, cognitive psychology and evolutionary linguistics within the ILM with humans, in order to develop a pioneering methodology for investigating the influence of language cultural transmission on the evolution of semantics (i.e., the co-evolution of language structure and semantics).

Our study was based on the assumption of the current experiment's continuity with Matthews et al.'s study (cf. § 5.1 for discussion). We designed and performed two analyses to demonstrate evolutionary changes in semantics that were driven by cultural transmission of L1 and L2.

In the first analysis we took into account differential category boundaries that evolved in L1 and L2 and hypothesised that perception of stimuli within and across evolved category boundaries should differ between the languages. If the differences in perception had been found, on the assumption of semantics grounded in perception (Barsalou, 1999), this would demonstrate that two distinct semantic systems evolved in L1 and L2. However, the non-significant result with respect to this analysis suggests that, although L1 and L2 cultural transmission resulted in the evolution of categorisation in Matthews et al.'s study, it might have led to too weak effects of within-category expansion and between-category compression that could not result in differential perception of these categories. It is relevant to emphasise, however, that this finding does not exclude the possibility that two distinct semantic systems evolved as a result of L1 and L2 transmission.

In fact, in our second analysis, we have demonstrated that learning languages that evolved two metrics of similarity – one that considered rotation relevant (i.e., L1) and the other that did not (i.e., L2) – resulted in differential perception of rotated and unrotated stimuli. This finding further implies that during language evolution, perception of features that are salient to human perceptual processing systems, such as rotation to the visual system, are the first ones to be emphasised, as in L1, or de-emphasised, as in L2, by evolving language labels. As a consequence, we can conclude that language evolution, by working together with human perceptual processing systems, can result in the evolution of qualitatively different semantic systems.

More importantly, however, the finding that cultural transmission of language can result in the evolution of different semantic systems suggests that not only does language carry information about its structure (e.g., compositionality – cf. Brighton et al., 2005), but also information about how its speakers should perceive and understand the world. The corollary of this suggestion should not remain underestimated as it indicates that, by shaping semantics, language evolution is committed to assuring mutual communicational intelligibility among members of given speech-communities. Taking as an example the languages from our study, we can suspect that speakers of L2, for who rotation is irrelevant, would be unable to understand why speakers of L1, for who rotation is important, think that rotated and unrotated shapes are different. Furthermore, it is worth emphasising that common understanding of the world is adaptive as it fosters closer relationships and empathy among speech community members.

Although previous experiments with the human ILM have demonstrated the co-evolution of language structure and semantics, they have not investigated changes in semantics directly by means of psychophysical measures.

For example, Kirby et al. (2008), using predefined stimuli as a semantics, demonstrated that during cultural transmission the semantics changes because parts of its elements are differentiated by gradual evolution of structure in strings that correspond to the elements. Furthermore, in Matthews et al.'s (in prep.) experiment, semantic systems differed across languages as a result of language transmission because we observed their differential organisation from scratch by means of formation of category boundaries and propagation of two metrics of similarity.

However in light of the significant finding in the current experiment, Matthews et al.'s results are reinforced because we can conclude that semantic systems in the minds of L1 and L2 speakers changed under the influence of respective language learning *because* we have observed differential perception of their elements. Therefore, the uniqueness of our approach in demonstrating co-evolution of language structure and semantics consists in the fact that we used a psychophysical measure of changes in semantics – that is – a measure of perceived similarity between objects in the meaning space.

Furthermore, from the perspective of cognitive psychology, it seems that the use of a psychophysical measure to demonstrate changes in semantics is more realistic than, for example, the use of abstract measures of structure, as in Kirby et al. (2008) and Matthews et al.'s (in prep) studies. This is because only the former measure is based on a cognitive psychological assumption that semantics of stimuli is grounded in sensorimotor representations that can be probed by measuring perception of these stimuli (cf. Chapter 1).

This point is relevant to research in evolutionary linguistics because it is crucial for findings in this field to be grounded in and supported by findings from other scientific disciplines that investigate human cognition. Thus, our study, as a step in this important direction, is highly relevant to the field of evolutionary linguistics.

Chapter 7 Conclusion

This study has offered an extension to the experiment by Matthews et al. (in prep.) who modelled evolution of categorisation of a continuous meaning space.

Matthews et al. demonstrated that language transmission within the ILM with human participants can lead to formation of category boundaries that are organised according to two metrics of similarity, one that recognises rotation of objects as relevant and the other that does not.

We have demonstrated that language transmission that propagates these two types of similarity metrics results in differential perception of rotated and unrotated objects.

This result is unprecedented in that it demonstrates the influence of language structure on semantics during language transmission by means of directly measuring cross-linguistic differences in perceived similarity between rotated and unrotated objects.

The current result has two major implications. Firstly, it demonstrates that not only can language evolution result in the emergence of different category systems (Matthews et al. in prep.) or compositionality (e.g., Kirby et al., 2008), but that it can also shape our conceptual system and the way we think about the world. This means that, by influencing semantics, language evolution acts as adaptive 'social glue' (Dijksterhuis, 2005:208) that fosters development of communicational intelligibility among all members of a given speech community and, as a result, improves within-community integrity.

Secondly, our finding supports linguistic relativity, according to which the language we speak influences our perceptual and conceptual systems. Although the theory of linguistic relativity is not new and has been vastly tested by psychologists in recent decades, there is a way whereby our result sheds new light on this approach. Linguistic relativity is usually tested with real natural languages considered synchronically, whereas the current result demonstrates cross-linguistic differences in perception as a result of learning languages evolved during 'speeded' evolution taking place in the laboratory.

References

Barsalou, L. W. (1999). Perceptual symbol systems. *The Behavioral and brain sciences*, *22*(4), 577-609; discussion 610-60.

Batali, J. (1998). Computational stimulations of the emergence of grammar. In J. Hurford, M. Studdert-Kennedy, & C. Knight, *Approaches to the Evolution of Language* (pp. 405-426). CUP.

Beale, J. M., & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition*, *57*(3), 217-39.

Belpaeme, T., & Bleys, J. (2005). Explaining Universal Color Categories Through a Constrained Acquisition Process. *Adaptive Behavior*, *13*(4), 293-310.

Bering, J. M. (2004). A critical review of the "enculturation hypothesis": the effects of human rearing on great ape social cognition. *Animal cognition*, *7*(4), 201-12.

Berlin, B., & Kay, P. (1969). *Basic colour terms: Their universality and evolution*. University of California Press.

Bickerton, D. (1990). *Language and Species*. University of Chicago Press.

Boyd, R., & Richerson, P. (1996). Why culture is common, but cultural evolution is rare. *Proceedings of the British Academy*, *88*, 77-93.

Brighton, H., Smith, K., & Kirby, S. (2005). Language as an evolutionary system. *Physics of Life Reviews*, *2*(3), 177-226.

Burns, E., & Ward, W. (1978). Categorical perception - phenomenon or epiphenomenon: Evidence from the perception of melodic musical intervals. *Journal of the Acoustical Society of America*, *63*, 456-468.

Cangelosi, A., & Harnad, S. (2002). The adaptive advantage of symbolic theft over sensorimotor toil: Grounding language in perceptual categories. *Evolution of Communication*, *4*(1), 117-142.

Cangelosi, A., Greco, A., & Harnad, S. (2000). From robotic toil to symbolic theft: grounding transfer from entry-level to higher-level categories. *Connection Science*, *12*(2), 143-162.

Chater, N., Reali, F., & Christiansen, M. H. (2009). Restrictions on biological adaptation in language evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(4), 1015-20.

Chomsky, N. (1995). *The Minimalist Program*. The Mitt Press.

Chomsky, N. (2002). *On nature and language*. CUP.

Christiansen, M. H. (2000). Using Artificial Language Learning to Study Language Evolution: Exploring the Emergence of Word Order Universals. In J. Dessalles & I. Ghadakpour, *Proceedings of the 3rd International Conference on the Evolution of Language* (pp. 45-48).

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *The Behavioral and brain sciences*, *31*(5), 489-508; discussion 509-58.

Christiansen, M. H., & Devlin, J. (1997). Recursive Inconsistencies are Hard to Learn: A Connectionist Perspective on Universal Word Order Correlations. In N. Mahwah, *Proceedings of the Ninetheenth Annual Conference of the Cognitive Science Society* (pp. 113-118). Lawrence Erlbaum Associates.

Christiansen, M. H., & Ellefson, M. R. (2002). Linguistic Adaptation without Linguistic Constraints: The Role of Sequential Learning in Language Evolution. In A. Wray, *The Transition to Language* (pp. 335-358). OUP.

Christiansen, M. H., Kelly, L., Shillcock, R., & Greenfield, K. (under review). Impaired artificial grammar learning in agrammatism.

Cornish, H. (2010). Investigating how cultural transmission leads to the appearance of design without a designer in human communication systems. *Interaction Studies*, *11*(1), 112-137.

Cornish, H., Tamariz, M., & Kirby, S. (2009). Complex Adaptive Systems and the Origins of Adaptive Structure: What Experiments Can Tell Us. *Language Learning*, *59*, 187-205.

Croft, W. (2000). *Explaining Language Change: an Evolutionary Approach*. England: Pearson Education Limited.

Deacon, T. (1997). *The Symbolic Species: The coevolution of language and human brain*. London Penguin.

Dijksterhuis, A. (2005). Why wer are social animals: The high road to imitation as social glue. In S. Hurley & N. Chater, *Perspectives on imitation: From neuroscience to social science, Volume 2* (pp. 207-220). MIT Press.

Dryer, M. S. (1992). The Greenbergian Word Order Correlations. *Language*, *68*(1), 81-138.

Ehret, G. (1987). Categorical perception of sound signals: Facts and hypotheses from animal studies. In S. Harnad, *Categorical Perception: The Groundwork of Cognition* (pp. 301-332). CUP.

Eimas, P., Siqueland, E., Jusczyk, P., & Vigorito, J. (1971). Speech perception by infants. *Science*, *171*, 303-306.

Elman, J. L. (1993). Learning and development in neural networks: the importance of starting small. *Cognition*, *48*(1), 71-99.

Estes, W. (1994). *Classification and Cognition*. OUP.

Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science*, *303*(5656), 377-80.

Francis, A., Ciocca, V., & Ng, B. (2003). On the (non)categorical perception of lexical tones. *Perception and Psychophysics, 65*, 1029-1044.

Franklin, A., Clifford, A., Williamson, E., & Davies, I. (2005). Color term knowledge does not affect categorical perception of color in toddlers. *Journal of experimental child psychology*, *90*(2), 114-41.

Franklin, A., Drivonikou, G. V., Bevis, L., Davies, I. R., Kay, P., Regier, T., et al. (2008a). Categorical perception of color is lateralized to the right hemisphere in infants, but to the left hemisphere in adults. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(9), 3221-5.

Franklin, a., Drivonikou, G. V., Clifford, A., Kay, P., Regier, T., Davies, I. R., et al. (2008b). Lateralization of categorical perception of color changes with color term acquisition. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(47), 18221-5.

Gallese, V., & Lakoff, G. (2005). The Brain's concepts: the role of the Sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, *22*(3), 455-479.

Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature*, *440*(7088), 1204-7.

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(2), 489-94.

Goldberg, D. (1989). *Genetic algorithms in search, optimization, and machine learning*. Reading, MA: Addison-Wesley.

Goldstone, R. (1994a). Influences of categorization on perceptual discrimination. *Journal of experimental psychology. General*, *123*(2), 178-200.

Goldstone, R. (1994b). The role of similarity in categorization: Providing a groundwork. *Cognition*, *52*, 125-157.

Goldstone, R. L., Lippa, Y., & Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition*, *78*, 27-43.

Goldstone, R., Day, S., & Son, J. (2010). Comparison. In B. Glatzeder, V. Goel, & A. von Muller, *On thinking: Volume II, towards a theory of thinking* (pp. 103-122). Springer Verlag.

Gordon, P., (2004). Numerical cognition without words: Evidence from Amazonia. *Science* 306, 496 (2004); 496-499.

Griffiths, T. L., & Kalish, M. (2007). Language Evolution by Iterated Learning with Bayesian Agents. *Cognitive Science*, *31*, 441-480.

Hahn, U., Chater, N., & Richardson, L. (2003). Similarity as transformation. *Cognition*, *87*, 1-32.

Harnad, S. (1987a). *Categorical perception: The groundwork of Cognition*. CUP.

Harnad, S. (1987b). Category induction and representation. In S. Harnad, *Categorical Perception: The Groundwork for Cognition* (pp. 535-566). CUP.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42*(1-3), 335-346.

Harnad, S. (2005). *To cognise is to categorise: Cognition is Categorisation*.

Harnad, S., Hanson, S., & Lubin, J. (1995). Learned categorical perception in neural nets: implications from sumbol grounding. In V. Honavar & L. Uhr, *Synbol Processors and Conneectionist Network Models in Aritificial Intelligence and Cognitive Modelling: Steps Toward Principled Integration* (pp. 191-206). New York: Acadmic Press.

Herman, L. (1984). Comprehension of sentences by bottlenosed dolphins. *Cognition*, *16*(2), 129-219.

Hodgetts, C. J., Hahn, U., & Chater, N. (2009). Transformation and alignment in similarity. *Cognition*, *113*(1), 62-79.

Hubel, D., & Wiesel, T. (1962). Perceptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, *160*, 106-154.

Kay, P., & Kempton, W. (1984). What is the Sapir-Whorf hypothesis? *American Anthropologist*, *86*, 65-78.

Kay, P., & Regier, T. (2007). Color naming universals: The case of Berinmo. *Cognition*, *102*, 289-298.

Kirby, S. (in prep.). The Evolution of Linguistic Replicators.

Kirby, S. (2000). Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. In C. Knight, *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form* (pp. 303-323). CUP.

Kirby, S. (2001). Spontaneous evolution of linguistic structure-an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation*, *5*(2), 102-110.

Kirby, S., & Hurford, J. (2002). The Emergence of Linguistic Structure: An Overview of the Iterated Learning Model. In A. Cangelosi & D. Parisi, *Simulating the Evolution of Language* (pp. 121-147). Springer.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(31), 10681-6.

Kirby, S., Smith, K., & Brighton, H. (2004). From UG to Universals: Linguistic adaptation through iterated learning. *Studies in Language*, *28*(3), 587-607.

Kroodsma, D., & Parker, L. (1977). Vocal virtuosity in the brwon thrasher. *The Auk*, *94*, 783-785.

Kuhl, P., & Miller, J. (1975). Speech perception by the chinchilla: Voice-voiceless distinction in alveolar plosive consonants. *Science*, *90*, 69-72.

Kuo, J. Y., & Sera, M. D. (2008). Classifier effects on human categorization: the role of shape classifiers in Mandarin Chinese. *Journal of East Asian Linguistics*, *18*(1), 1-19.

Ladd, D., & Morton, R. (1997). The perception of intonational emphasis: Continuous or categorical? *Journal of Phonetics*, *25*, 313-342.

Lamb, T. D., Collin, S. P., & Pugh, E. N. (2007). Evolution of the vertebrate eye: opsins, photoreceptors, retina and eye cup. *Nature reviews. Neuroscience*, *8*(12), 960-76.

Levenstein, V. (1966). Binary codes capable of correcting deletions, insertions and reversals.

Levin, D., & Beale, J. (2000). Categorical perception occurs in newly learned faces, other-race faces, and inverted faces. *Percept Psychophysics*, *62*, 386-401.

Levinson, S. C. (1996). Language and Space. *Annual Review of Anthropology*, *25*(1), 353-382.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358-368.

Lindsey, D., & Brown, A. (2006). Universality of color names. *Proceedings of the National Academy of Sciences*, *103*, 16609-16613.

Livingston, K. R., Andrews, J. K., & Harnad, S. (1998). Categorical perception effects induced by category learning. *Journal of experimental psychology. Learning, memory, and cognition*, *24*(3), 732-53.

Lucy, J. (1997). The linguistics of 'color'. In C. Hardin & L. Maffi, *Color categories in thought and language* (pp. 320-346). CUP.

Lupyan, G. (2008). The conceptual grouping effect: categories matter (and named categories matter more). *Cognition*, *108*(2), 566-77.

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: redundant labels facilitate learning of novel categories. *Psychological science*, *18*(12), 1077-83.

Lupyan, G., Thompson-Schill, S. L., & Swingley, D. (2010). Conceptual Penetration of Visual Processing. *Psychological Science*, *21*(5), 682-691.

Matthews, C., Kirby, S., & Cornish, H. (in prep.). The emergence of categorisation: Language Transmission in an Iterated Learning Model using a continuous meaning space.

McClelland, J., & Rumelhart, D. (1986). On learning the past tenses of English verbs. In J. L. McClelland, D. Rumelhar, & P. R. Group, *Parallel Distributed Processing, Volume 2* (pp. 225-258). The MIT Press.

Merker, B., & Okanoya, K. (2007). The Natural History of Human Language: Bridging the Gaps without Magic. In C. Lyon, C. Nehaniv, & A. Cangelosi, *Emergence of Communication and Language* (pp. 403-420). Springer.

Mesoudi, A., & Whiten, A. (2008). Review. The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *363*(1509), 3489-501.

Nakisa, R., & Plunkett, K. (1998). Evolution of a rapidly learned representation for speech. *Language and Cognitive Processes*, *13*, 105-127.

Paradiso, M. a. (2002). Perceptual and neuronal correspondence in primary visual cortex. *Current opinion in neurobiology*, *12*(2), 155-61. Parisi, D. (1997). An Aritficial Life approach to language. *Mind and Language*, *59*, 121-146.

Payne, K. (2000). The progressively changing songs of humpback whales: A window on the creative process ina wild animal. In N. Wallin, B. Merker, & S. Brown, *The Origins of Music* (pp. 135-150). The MIT Press.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, *13*(4), 707-784.

Pinker, S., & Prince, A. (1989). Rules and connections in human language. In R. Morris, *Parallel Distributed Processing* (pp. 307-318). OUP.

Price, E. E., Caldwell, C. a., & Whiten, A. (2010). Comparative cultural cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(1), 23-31.

Pullum, G. K., & Scholz, B. C. (2002). Empirical assessment of stimulus poverty arguments. *The Linguistic Review*, *19*(1-2), 9-50.

Reali, F., & Christiansen, M. H. (2009). Sequential learning and the interaction between biological and linguistic adaptation in language evolution. *Interaction Studies*, *10*, 5-30.

Regier, T., & Kay, P. (2009). Language, thought, and color: Whorf was half right. *Trends in cognitive sciences*, *13*(10), 439-46.

Roberson, D., Davidoff, J., & Braisby, N. (1999). Similarity and categorisation : neuropsychological evidence for a dissociation in explicit categorisation tasks. *Cognition*, *71*, 1-42.

Roberson, D., Davidoff, J., & Shapiro, L. (2002). Squaring the Circle: The Cultural Relativity of 'Good' Shape. *Journal of Cognition and Culture*, *2*(1), 29-51.

Roberson, D., Davidoff, J., Davies, I. R., & Shapiro, L. R. (2005). Color categories: evidence for the cultural relativity hypothesis. *Cognitive psychology*, *50*(4), 378-411.

Roberson, D., Davies, I., & Davidoff, J. (2000). Color categories are not universal: Replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General*, *3*, 369-398.

Rosch Heider, E., & Olivier, D. (1972). The structure of the color space in naming and memory for two languages. *Cognitive Psychology*, *3*, 337-354.

Rosch, E. (1973). Natural categories. *Cognitive Psychology*, *4*, 328-350.

Saunders, B., & van Brakel, J. (1997). Are there non-trivial constraints on colour categorisation. *Behavioural and Brain Sciences*, *20*, 167-178.

Savage-Rumbaugh, E. S., Murphy, J., Sevcik, R. a., Brakke, K. E., Williams, S. L., Rumbaugh, D. M., et al. (1993). Language comprehension in ape and child. *Monographs of the Society for Research in Child Development*, *58*(3-4), 1-222.

Schusterman, R., & Gisiner, R. (1988). Artificial Language Comprehension in Dolphins and Sea Lions: the Essential Cognitive Skills. *The Psychological Record*, *38*, 311-348.

Searle, JR (1982) The Chinese room revisited. *Behavioral and Brain Sciences*,*5*:345-348

Seyfarth, R., Cheney, D., & Marler, P. (1980). Vervet monkey alarm calls: Semantic communication in a free-ranging primate. *Animal Behaviour*, *28*(4), 1070-1094.

Shepard, R. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, *22*, 325-345.

Shepard, R. N. (1974). Representation of structure in similarity data: Problems and prospects. *Psychometrika*, *39*(4), 373-421.

Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: a similarity-based model. *Journal of experimental psychology. General*, *133*(2), 166-88.

Smith, E. (1989). Concepts and Induction. In M. Posner, *Foundations of Cognitive Science* (pp. 501-526). The MIT Press.

Smith, K. (2002). The cultural evolution of communication in a population of neural networks. *Connection Science*, *14*(1), 65-84.

Smith, K. (2004). The evolution of vocabulary. *Journal of theoretical biology*, *228*(1), 127-42.

Snowdon, C. (1987). A naturalistic view of categorical perception. In S. Harnad, *Categorical Perception: The Groundwork of Cognition* (pp. 332-355). CUP.

Steels, L., & Belpaeme, T. (2005). Coordinating perceptually grounded categories through language: a case study for colour. *The Behavioral and brain sciences*, *28*(4), 469-529.

Stevens, K., Liberman, A., Studdert-Kennedy, M., & Ohman, S. (1969). Crosslanguage study of vowel perception. *Language and Speech*, *12*, 1-23.

Tallerman, M. (2007). Did our ancestors speak a holistic protolanguage? *Lingua*, *117*(3), 579-604.

Terrace, H., Petitto, L., Sanders, R., & Bever, T. (1979). Can and Ape Create a Sentence? *Science*, *206*, 891-901.

Tomasello, M. (1999). the Human Adaptation for Culture. *Annual Review of Anthropology*, *28*(1), 509-529.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327-352.

Whorf, B. L. (1956). *Language, thought and reality: Selected Writings of Benjamin Lee Worf*. (J. Carroll). The MIT Press.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic bulletin & review*, *9*(4), 625-36.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., Boroditsky, L., et al. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(19), 7780-5.

Wray, A. (1998). Protolanguage as a holistic system for social interaction. *Language and Communication*, *18*(1), 47-67.

Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition*, *85*(3), 223-50.

Yoshida, H., & Smith, L. B. (2005). Linguistic cues enhance the learning of perceptual cues. *Psychological science : a journal of the American Psychological Society / APS*, *16*(2), 90-5.

Özgen, E., & Davies, I. R. (1998). Turkish color terms: tests of Berlin and Kay's theory of color universals and linguistic relativity. *Linguistics*, *36*(5), 919-956.

Appendices

Appendix A1: Categorisations of the 100-element meaning space generated by a perfect learner acquiring L1

Appendix A2: Categorisations of the 100-element meaning space generated by a perfect learner acquiring L2

Appendix B: Experimental Instructions

Appendix C: The list of stimuli from the similarity judgement task

Appendix A1

1 MANI	2 MANI	3 MANI	4 MANI	5 NIKIHE	6 NIKIHE	7 NIKIHE	8 NIKIHEHE	9 NIKIHE	10 NIKIHE
11 MANI	12 MANI	13 PANI	14 NIKIHEHE	15 NIKIHEHE	16 NIKIHE	17 NIKIHEHE	18 NIKIHEHE	19 NIKIHE	20 NIKIHE
21 PANI	22 PANI	23 PANI	24 PANI	25 NIKIHEHE	26 NIKIHEHE	27 NIKIHEHE	28 NIKIHE	29 NIKIHE	30 NIKIHE
31 PANI	32 PANI	33 PANI	34 NIKIHEHE	35 NIKIHEHE	36 NIKIHEHE	37 NIKIHE	38 NIKIHE	39 NIKIHE	40 NIKIHE
41 PANI	42 PANI	43 PANI	44 NIKIHEHE	45 NIKIHEHE	46 NIKIHEHE	47 NIKIHE	48 NIKIHE	49 NIKIHE	50 NIKIHE
51 PANI	52 PANI	53 PANI	54 NIKIHEHE	55 NIKIHEHE	56 PANI	57 PANI	58 NIKIHEHE	59 NIKIHEHE	60 NIKIHE
61 PANI	62 PANI	63 PANI	64 PANI	65 PANI	66 PANI	67 PANI	68 NIKIHEHE	69 NIKIHEHE	70 MANI
71 PANI	72 PANI	73 PANI	74 PANI	75 PANI	76 PANI	77 MANI	78 MANI	79 MANI	80 MANI
81 PANI	82 PANI	83 PANI	84 PANI	85 PANI	86 MANI	87 MANI	88 MANI	89 MANI	90 MANI
91 PANI	92 PANI	93 PANI	94 PANI	95 PANI	96 PANI	97 MANI	98 MANI	99 MANI	100 MANI

Appendix A2

1 MAUKIKI	2 MAUKIKI	3 MAUHIHI	4 MAUHIHI	5 MAUMAUAGE	6 MAUMAUAGE	7 MAUMAUAGE	8 MAUMAG	9 MAUMAG	10 MAUMAG
11 MAUKIKI	12 HIHI	13 HIHI	14 MAUHIHI	15 MAUHIHI	16 MAUMAUAGE	17 MAUMAGE	18 MAUMAG	19 MAUMAG	20 MAUMAG
21 MAUKIKI	22 MAUHIHI	23 HIHI	24 MAUKIHI	25 MAUHIHI	26 MAUMAGE	27 MAUMAG	28 MAUMAG	39 MAUMAG	30 MAUMAG
31 MAUHIHI	32 MAUHIHI	33 MAUKIHI	34 MAUKIHI	35 MAUKIHI	36 MAUKIHI	37 MAUMAUAGE	38 MAUMAUAGE	39 MAUMAUAGE	40 MAUMAUAGE
41 MAUKIHI	42 MAUKIHI	43 MAUKIHI	44 MAUKIHI	45 MAUKIHI	46 MAUKIHI	47 MAUKIHI	48 MAUKIHI	49 MAUKIHI	50 MAUKIHI
51 MAUKIHI	52 MAUKIHI	53 MAUKIHI	54 MAUKIHI	55 MAUKIHI	56 MAUKIHI	57 MAUKIHI	58 MAUKIHI	59 MAUKIHI	60 MAUKIHI
61 MAUMAUAGE	62 MAUMAUAGE	63 MAUMAUAGE	64 MAUMAUAGE	65 MAUKIHI	66 MAUKIHI	67 MAUKIHI	68 MAUKIHI	69 MAUHIHI	70 MAUHIHI
71 MAUMAG	72 MAUMAG	73 MAUMAG	74 MAUMAG	75 MAUMAGE	76 MAUHIHI	77 MAUKIHI	78 HIHI	79 MAUHIHI	80 MAUKIKI
81 MAUMAG	82 MAUMAG	83 MAUMAG	84 MAUMAGE	85 MAUMAUAGE	86 MAUHIHI	87 MAUHIHI	88 HIHI	89 HIHI	90 MAUKIKI
91 MAUMAG	92 MAUMAG	93 MAUMAG	94 MAUMAUAGE	95 MAUMAUAGE	96 MAUMAUAGE	97 MAUHIHI	98 MAUHIHI	99 MAUKIKI	100 MAUKIKI

Appendix B

Welcome to Alpha-3–6a in a galaxy far, far away.

We have encountered an intelligent alien life form with its own form of language. You must try to learn this language as best you can.

Don’t worry if you feel overwhelmed—the alien knows that this is a difficult task for you to master, and it will do its best to understand everything that you say.

(When you are ready to continue, press ENTER).

You will see a series of pictures and the way in which the alien would describe those pictures.

Every now and then the alien will test your knowledge of the language by showing you a picture without any description. Simply write what you think the correct label is and press (ENTER).

The alien will also want to know if you see pictures how he does.

So at some point it will ask you to judge the similarity between pairs of pictures.

DON'T WORRY IF YOU FEEL YOU HAVE NOT YET MASTERED THE LANGUAGE!

The most important thing is to maintain good relations with the aliens and give it your best shot.

ALWAYS GIVE AN ANSWER. That way the aliens will know you are trying.

They will go out of their way to try to understand everything you say and they are very patient.

You will be given a break every few minutes.

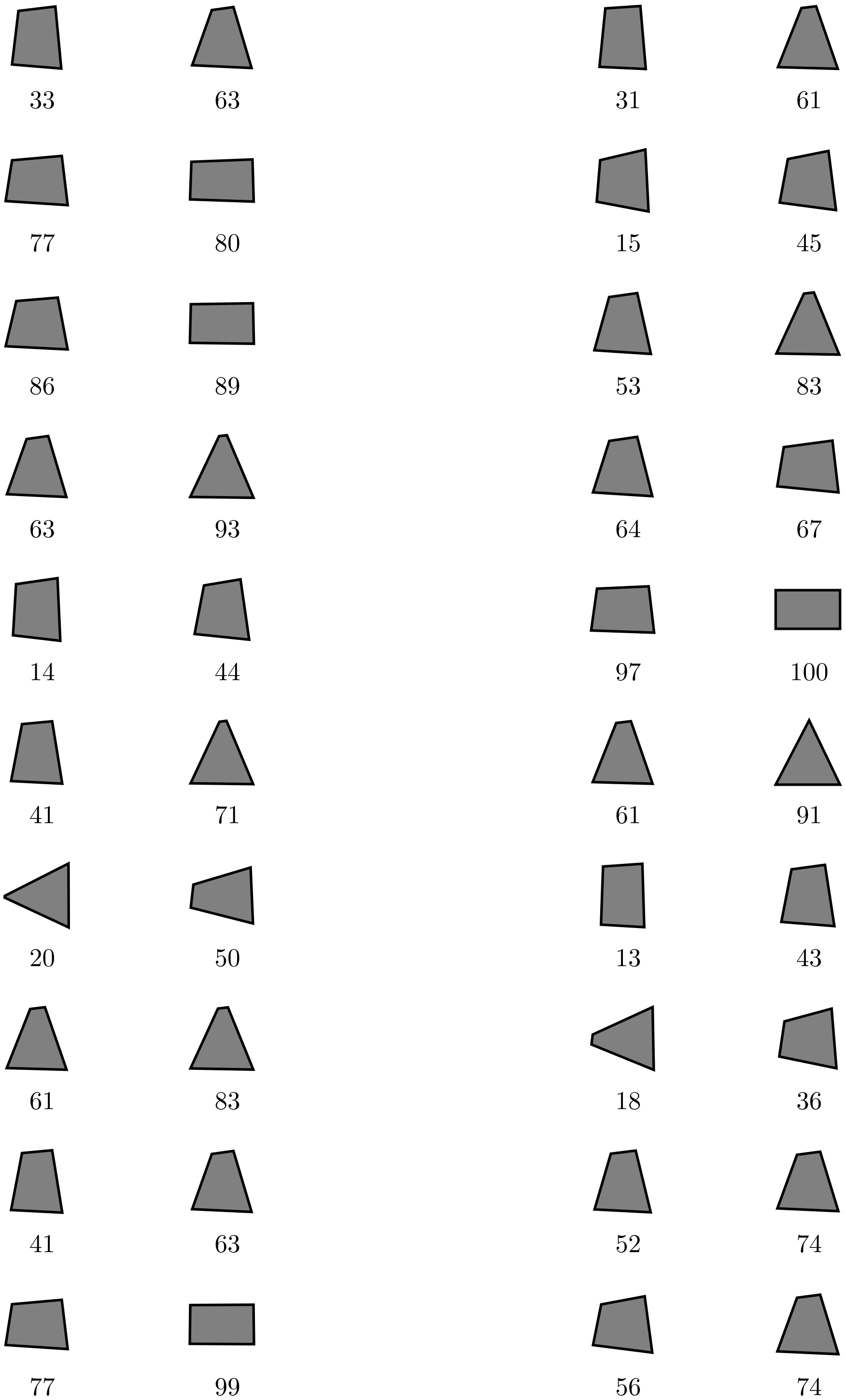
If you have any questions please ask the experimenter now.

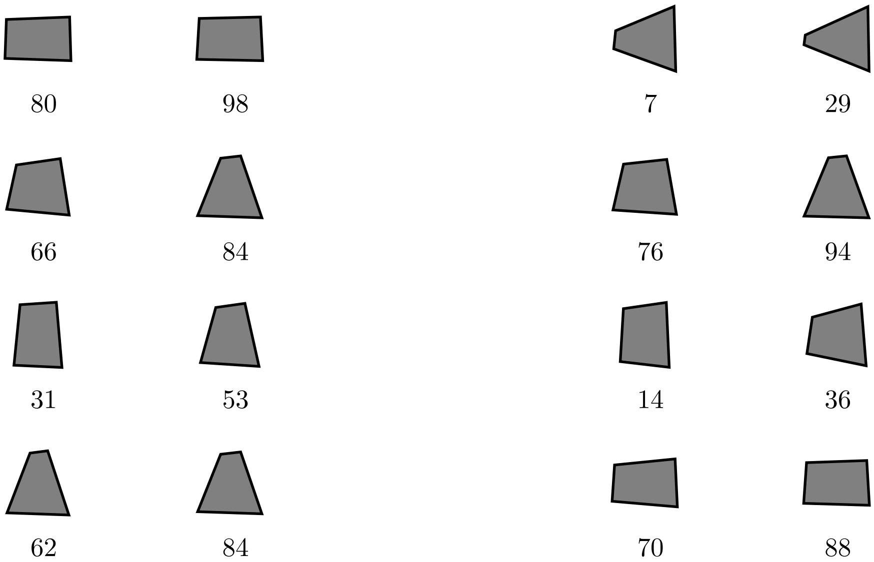
GOOD LUCK!

(press ENTER to start the tuition)

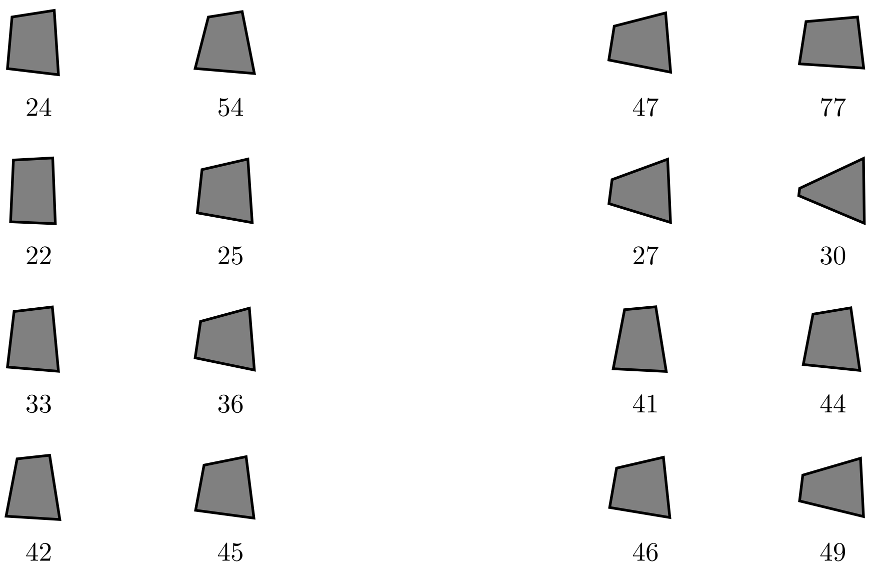
Appendix C

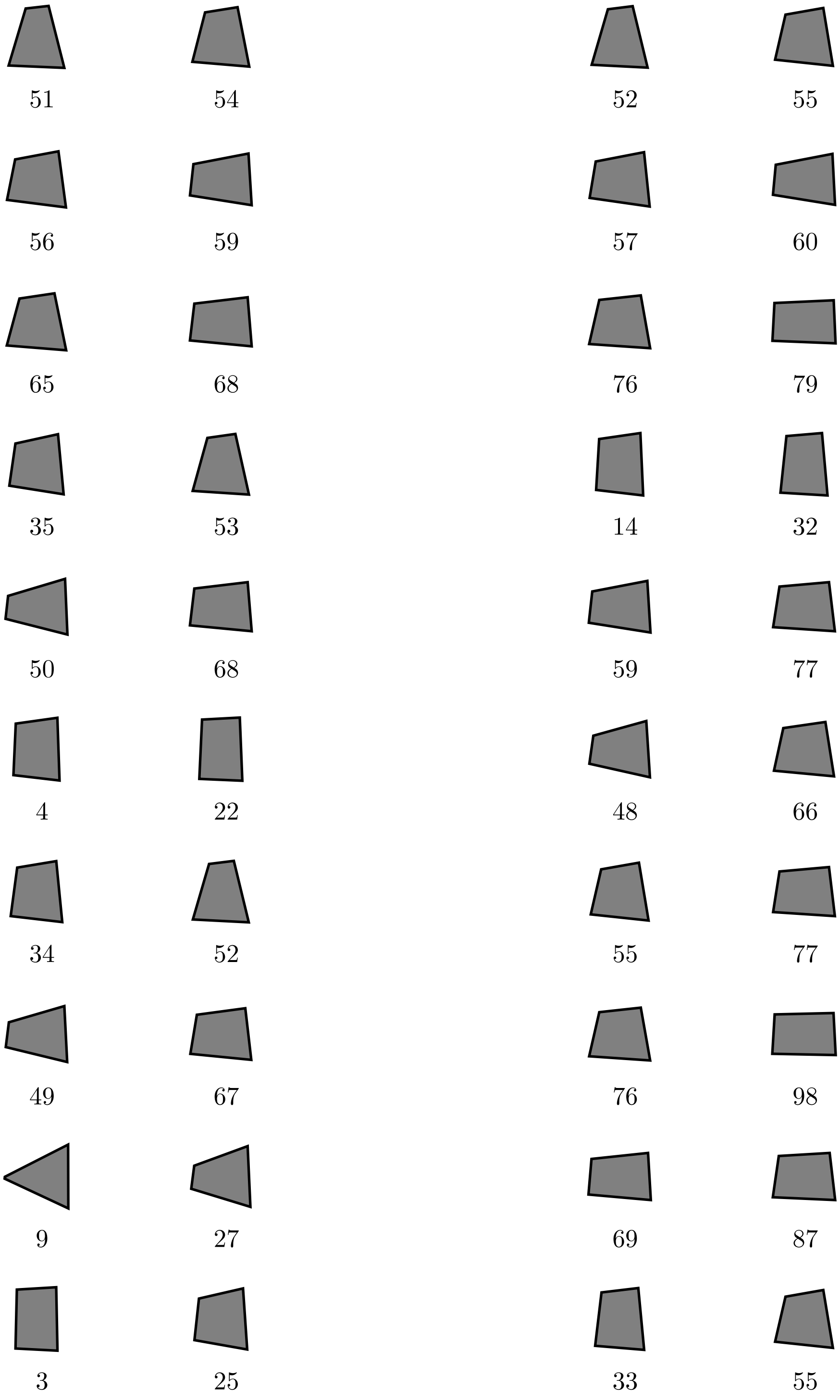
Shape pairs categorized as Same in L1 and Different in L2



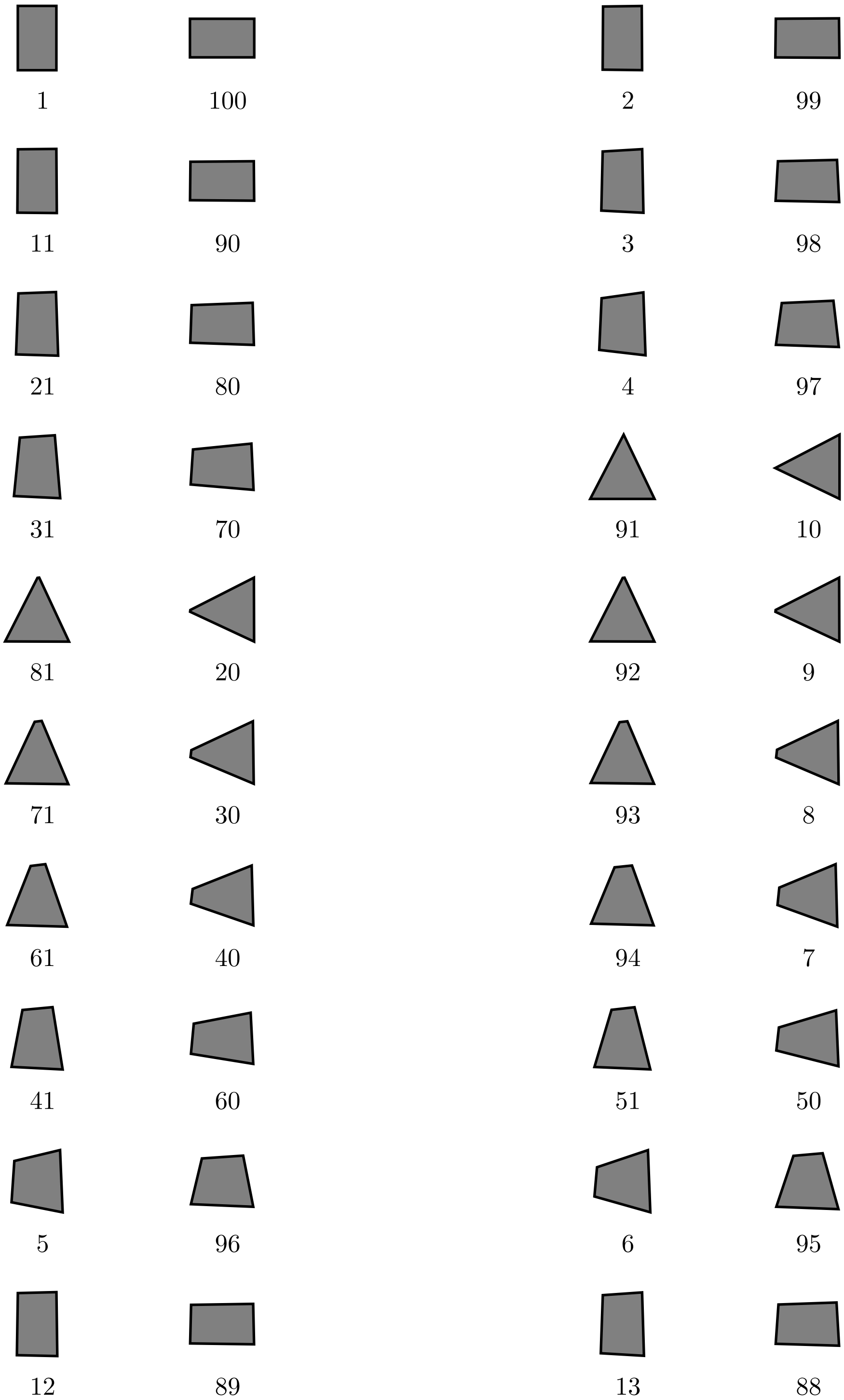


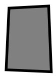
Shape pairs categorized as Different in L1 and Same in L2



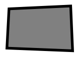

51 54
56 59
65 68
35 53
50 68
4 22
34 52
49 67
9 27
3 25
52 55
57 60
76 79
14 32
59 77
48 66
55 77
76 98
69 87
33 55

Pairs with rotated and unrotated shapes

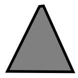




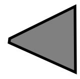
22



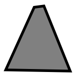
79



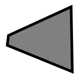
82



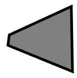
19



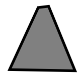
72



29



18



83